Erasmus University Rotterdam
Erasmus School of Economics

Master Thesis Econometrics and Management Science: Quantitative Finance

Enhancing Flood Risk Assessment and Reinsurance Strategies in the Netherlands: A Quantile Estimation Approach

Author:
Vincent de Quelerij (464628)
External Supervisor:
Msc. S.C.F.W. van Eekelen

Supervisor:
Prof. Dr. D.J.C. van Dijk
Second assessor:
Prof. Dr. C. Zhou

1 November 2023

Abstract

Climate change exacerbates flood risk within the Netherlands, a country that is already situated below sea level. Furthermore, the problem of data scarcity results into low awareness and clarity regarding damages as a consequence of floods. This research proposes and demonstrates a novel image recognition algorithm to overcome data limitations and extract damage proxies from flood scenario maps for fluvial and coastal floods. We perform a quantile fitting analysis to obtain a loss distribution, which forms the basis for assessing the viability of CAT bonds as a potential reinsurance pool. Additionally, we show how increasing awareness of risk exposure yields benefits not only applicable to insurance company portfolio decision making but also to national policy considerations. Moving forward, we find that CAT bonds deem a reasonable option as a reinsurance pool within the Netherlands and show the optimal strategy for issuing CAT bonds from the insurers' perspective for a vast range of CAT bond parameters, by applying both analytical and numerical optimization methods.

Acknowledgements

I would like to express my gratitude to Prof. Dr. D.J.C. van Dijk for his support, supervision and guidance throughout the process of writing my master's thesis. His invaluable knowledge and feedback elevated the quality and findings of this research.

Additionally, I want to thank S.C.F.W. van Eekelen from PricewaterhouseCoopers (PwC). He provided me with valuable insights, perspectives and suggestions, while consistently monitoring my project's advancement.

Contents

1 Introduction

The Netherlands has a vast history with floods, mainly due to the fact that the country is under sea level. The threat of climate change within recent years exacerbates the flood risk even more. Hence, the Netherlands serves as a key example to address and examine the probability and severity of potential damages and how insurance companies and governments should manage the increasing risks. Scarcity of historic data poses as one of the major challenges when assessing costs of flood risk. Largely sophisticated models exist on a global level to provide an indication of the probability and severity of pluvial, fluvial and coastal flood risks. However, the translation from such risks to associated costs proves difficult up to today (Dutch Association of Insurers, 2020).

This thesis contributes to existing literature by suggesting a novel methodology to overcome both the problem of data scarcity and lack of clarity regarding costs due to flood damages. Additionally, we explore the viability of CAT bonds as a reinsurance pool in the Netherlands. We believe that this is a pivotal step in the current dialogue about insurance clarity in the Netherlands (Dutch Association of Insurers, 2022). To overcome data scarcity, we consult LIWO, Watermanagementcenter Nederland (2023), who present four very well established flood scenario maps within the Netherlands, all corresponding to a particular probability of occurrence, which may be interpreted as a quantile of some distribution. Additionally, the scenario maps distinguish between different flood depths. Note that these flood maps exclude pluvial floods that arise due to precipitation. This thesis translates the flood scenario maps towards loss distributions. We focus in particular on damages regarding residential buildings. This subcategory of damages is highly relevant for insurance companies. Additionally, data on the value of residential buildings is widely and publicly available within the Netherlands. This evokes the first research question which formulates as follows:

RQ1: *How can we extract and quantify damages to residential buildings from existing flood risk scenarios within the Netherlands?*

This thesis attempts to answer this question in two steps. First, we provide a flexible procedure for translating existing flood scenario maps towards associated damage proxies. This thesis proposes a novel image recognition approach to overcome data limitation of flood risk. We perform the procedure on South-Holland - a province in the Netherlands - and translate four flood scenario maps to damage proxies for damages to residential buildings. Additionally, we provide a sensitivity analysis.

As a second step, we obtain a more comprehensive assessment of the costs by consulting a quantile fitting analysis. In this analysis, we fit four heavy-tailed distributions to the extracted damage proxies corresponding to the different quantiles associated with the flood risk scenario maps. We compare the extracted damage proxy quantiles to the theoretical quantiles of these distributions by using different error metrics and choose the best distribution fit. This allows us to simulate yearly as well as multi-yearly losses from a distribution, not only restricted to the four flood map scenario quantiles. Furthermore, we illustrate how insurance companies benefit from knowing potential losses within their client portfolio.

In the Netherlands, the Dutch Association of Insurers (DAI) currently works on a plan regarding compensation for flooding damage caused by major rivers or the sea (Dutch Association of Insurers, 2022). They propose a system that uses a public-private international reinsurance pool to obtain wide risk sharing of the damages. The reason for this new plan is that current rules and regulation on flood damages are not explicit and clear for insurers to get a concrete overview of the costs corresponding to flood risk.

Currently, other countries make use of Catastrophe (CAT) bonds to partially reallocate risk towards investors. To our current knowledge, there does not exist any sort of CAT bond within the Netherlands. We examine whether this is a potential option within the Netherlands by attempting to answer the second research question:

RQ2: *How can insurance companies use CAT bonds as a reinsurance pool for flood risk within the Netherlands?*

In this case, we explore zero-coupon CAT bonds for which investors generally receives a premium after the bond expires. In the case of damages exceeding an attachment point, the investor only receives back a fraction of that payment. We consider the insurers' perspective and investigate if it is in their best interest to reallocate risk through CAT bonds by using a mean-variance hedge to determine the optimal amount of CAT bonds for an insurance company to issue. We consider a variety of risk aversion and bond parameters. We perform mean-variance optimization both analytically and by simulation, where we simulate hypothetical losses from the fitted distribution and estimate the payoff from the insurers' perspective. If the optimal amount of CAT bonds to be issued is zero, this indicates that CAT bonds prove no viable option as a reinsurance pool. We extensively investigate the change in the optimal amount of CAT bonds to be issued when changing parameters individually or combined.

This thesis obtains damage proxies for each flood scenario quantile map, ranging from 44

million up to 88 billion euros for South-Holland. The degree of these numbers confirms the importance of clarity about the potential damages due to flooding. Next, We conduct a quantile fitting analysis to obtain a corresponding loss distribution, where the generalized extreme value (GEV) distribution performs best. We optimize the distribution parameters such that the theoretical quantile values best align with the extracted damage proxies corresponding to their respective quantiles. This optimization is done by using four error metrics. For this application, the mean squared error (MSE) and mean squared percentage error (MSPE) provide the most accurate distribution fits compared to the mean absolute error (MAE) and mean absolute percentage error (MAPE).

Next, we find that from the insurers' perspective, CAT bonds are generally a valid option as a reinsurance pool for multiple degrees of risk aversion. We find the optimal amount of CAT bonds to be issued to be above zero for a wide range of bond parameters. Furthermore, we show how multiple bond parameters affect the optimal strategy for insurers when issuing CAT bonds. For example, we find that higher risk aversion leads to more CAT bonds to issue to reallocate more risk. The effects of the payout fraction in the case of a catastrophic event rely on the premium paid by the insurer. Lastly, the attachment point should be considered carefully as it has a concave parabolic effect on the optimal amount of CAT bonds to issue.

The remainder of this thesis is structured as follows. Section 2 contains an overview of relevant literature to help us answer the research questions. Secondly, Section 3 describes and visualizes the data and provides some summary statistics. Next, Section 4 elaborates on the methodology and gives some necessary derivations. Section 5 shows in-depth results for every described method and provides the reader with insights as well as a sensitivity analysis. Section 6 concludes. Lastly, Section 7 addresses limitations and possible further research.

2 Literature Review

2.1 Heavy Tailed Distributions

After obtaining proxies for flood losses for different quantiles, we aim to convert these quantiles estimates to an actual distribution from which we draw losses. Recent studies use quantile regression methods to assess tail risks of various macroeconomic and financial variables. Since the flood scenario maps correspond to their respective quantiles, this is an interesting field to further examine.

Adrian et al. (2019) estimate a quantile regression of typically four chosen quantiles, which is equal to the number of quantiles we have available for this thesis. Mitchell et al. (2022) discuss

robustness regarding the choice of the amount of quantiles used for fitting a distribution. They find supporting results for the rather arbitrary choice of four quantiles and state that this is often sufficient in practice.

As a second step, Adrian et al. (2019) fit a skewed-t density function to the quantile estimates by minimizing the Eucledian distance (or ℓ_2 norm) between the empirical quantiles and the theoretical quantiles of the chosen density. Note that this method goes beyond only the skewed-t density and other distributions can be used. In the application for flood risk, other distributions may be more suitable. Due to the heavy-tailed nature of damages due to flood risks, we aim to fit other heavy-tailed distributions such as the generalized pareto distribution (GPD) and GEV distribution to quantile data (Hattermann et al., 2014; Morrison and Smith, 2002).

2.2 Theoretical Framework CAT Bonds

The fitted distribution allows for simulating hypothetical losses, which we use to research the possibility of CAT bonds as a reinsurance pool. CAT bonds still remain absent in the Netherlands, while other countries make use of these financial instruments for different purposes. As final part of this thesis, we conduct research on various implementations of CAT bonds within the Netherlands. Luckily, literature is rich on CAT bonds and its different possible structures (Zonggang and Ma, 2013; Stupfler and Yang, 2017; Lee and Yu, 2007).

Zonggang and Ma (2013) show that a general scheme for CAT bonds works as follows. The insurer or sponsor wishes to transfer risk towards one or multiple investors that in return demand higher expected returns for accepting the risk. Insurers create a special purpose vehicle (SPV) that serves to isolate the specific risk that is to be reinsured. The insurer pays a premium to the SPV in exchange for a pre-defined amount of coverage in the case of a catastrophic event. On the investor side, the SPV receives principals from the investors and after the bond expires, pays back the principal plus a premium appropriate for the risk. The SPV securely invests the received capital into safe short-term securities such as Treasury bonds that are held in a trust account. The almost riskless returns from these securities are then swapped for more fluctuating returns that depend on the London Interbank Offered Rate (LIBOR) via a highly rated swap counter party. The use of swaps allows for both the investors and insurer to mitigate interest rate risk and default risk. For simplification, we assume no interest or default risk in this thesis. Visualization of this framework shows in Figure 1.

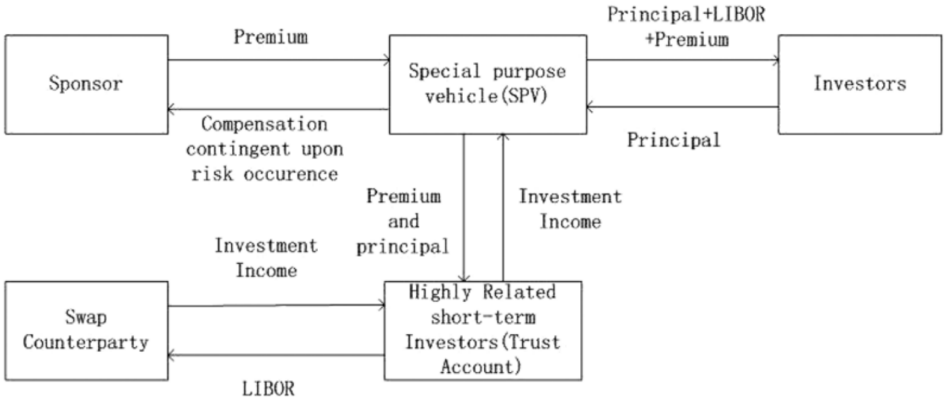


Figure 1: Structure of CAT bond transactions. This figure is gathered from Zonggang and Ma (2013).

In the scenario of a catastrophic event, the investors receive their principal back only partially or not at all, depending on the CAT bond contract. The SPV in turn pays the insurer the agreed coverage amount.

3 Data

3.1 Flood Scenario Maps

The data consists of four flood maps of the Netherlands with scenarios that correspond to the following flood probability quantiles per year: 0.9, 0.99, 0.999 and 0.9999. All flood scenario maps are retrieved from LIWO, Watermanagementcenter Nederland (2023). The underlying map contains very detailed information on positions and areas of buildings, roads, water etcetera, all distinguished by different colors. Every quantile corresponds to its own flood scenario, an overarching layer that is placed on the underlying map, containing a pixels with colors corresponding to different maximum flood depths. These flood depths correspond to flood occurrences associated with the probabilities. The grid size of the pixels varies depending on whether a flood is fluvial (smaller) or coastal (larger). Figure 2 shows a fully zoomed out 1 in 100 year flood scenario map of the Netherlands, corresponding to the 0.99 quantile. Observe that the map contains the option to include municipality borders, which allows us to assess flood damage for municipalities individually. An important note is that every scenario map has the underlying assumption that the probability of such a scenario remains the same over the years due to policies on improving the dikes before 2050. The main purpose of Figures 2 is to give the reader a general feeling of the visual data that we handle in the remainder of this thesis.

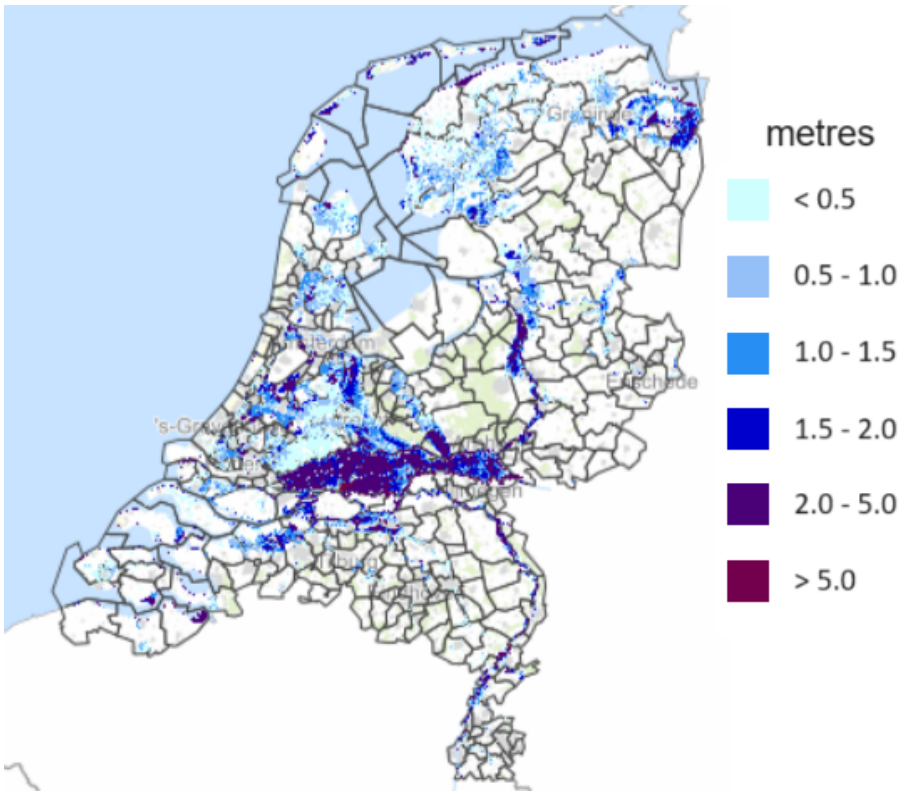


Figure 2: Visual data of the 1 in 100 year flood risk scenario in the Netherlands. The scenario shows the maximum depth of the flood for different scenarios in meters.

Hereafter, we use zoom 6 - a setting - of the flood scenario map for the data quantifying process. This zoom provides for a solid distinction between pixels that do or do not contain buildings and flooded pixels corresponding to different maximum flood depths, while keeping data preparation time to a minimum. The scenario maps are not available for download to the best of our knowledge. In case that the map can be downloaded in high resolution, we of course should consider the maximum zoom for the most accurate results. We later perform a sensitivity analysis to investigate the effects of adjusting the zoom. We further elaborate on the data handling process in Section 4.1.

Due to no available download, we work with screenshots where we prepare six images for every municipality that perfectly align with each other to the pixel. Consider the municipality Gouda as an example in Figure 3. Figure 3a includes no underlying map and only the municipality borders to determine which pixels lie within the municipality. Figure 3b serves to determine which pixels correspond to buildings. Finally, Figures 3c-3f correspond to the four flood scenario maps to determine which building pixels are at risk for flooding.

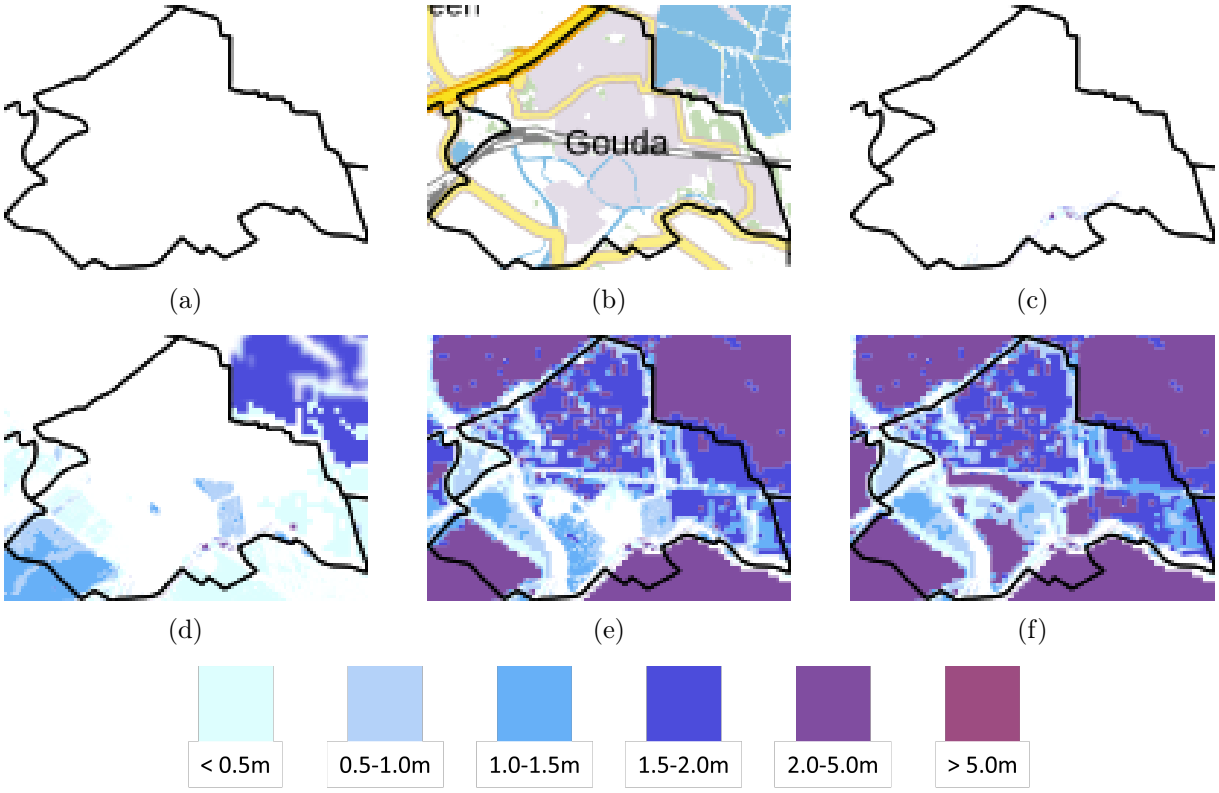


Figure 3: Six exactly aligned screenshots of the municipality Gouda. Images include municipality borders, a colored underlying map that includes buildings in light grey. Additionally, four flood scenarios with different maximum flood depths show corresponding to 1 in 10, 100, 1000 and 10000 year floods.

3.2 Buildings

To assist the extraction of damage proxies from the flood scenario maps, we use two additional data sets. First, we use data from Statistics Netherlands (2023a) that consists of the total number of buildings per municipality in the Netherlands in January, 2023. We use this data set in combination with the flood scenario maps to translate the percentages of flooded buildings, based on pixel color codes, to actual numbers of buildings at risk. The 2023 data set contains 342 municipalities.

A third data set connects the total number of buildings to their respective WOZ value for every individual municipality. The data is retrieved from Statistics Netherlands (2023b) and contains average WOZ values in euros per municipality at January, 2023. We use WOZ values as a proxy for the total possible damage costs to a building. The market value is a good or arguably better alternative to consider for such a proxy. However, data on WOZ values is easily accessible and up to date for every municipality, while market values are harder to retrieve publicly. This last data set for WOZ values is crucial for converting the four flood scenarios and number of buildings at risk to a proxy of actual potential damages in euros. Table 1 shows some summary

statistics for both the amount of buildings per municipality and its average WOZ values. Observe that damages can differ strongly between municipalities due to large differences in both buildings and WOZ values.

Table 1: Summary statistics of the total amount of buildings and WOZ values for each municipality within the Netherlands

	Total	Mean	Maximum	Minimum
Buildings	9213833	26707	507344	733
WOZ values (€)		378614	985000	192000

The flood scenario maps assign the same color code to every building, so we are not able to differentiate between residential and other buildings. Therefore, we make the assumption that every building is a residential building. Pixels may contain one or multiple buildings, depending on the zoom. This may results into some inaccuracies in the estimation of damage proxies. However, this thesis mainly focuses on providing a general procedure of extracting damage proxies from flood scenario maps. We aim to approach the potential damages and are bond to make some assumptions in the process due to time limitations.

3.3 Depth-Damage Curves

Available flood maps of the Netherlands include the maximum flood depth in meters to give additional insights about the severity of the flood. Since higher flood depths result into higher damage costs and vice versa, this should be taken into account when creating a proxy for damage costs.

A common way to estimate direct flood damages is through depth-damage curves (Huizinga et al., 2017). Such curves distinguish flood damages at specific depth ranges, often based on historic events or expert judgements. Huizinga et al. (2017) present depth-damage curves for different countries and damage categories. Figure 4 shows depth-damage curves for residential building damages for countries located in Europe, including the Netherlands. The damage factor is equal to the percentage of the sum insured at risk. We use the damage factors for the Netherlands in this thesis to access damages for different flood depths. Based on this figure, a Dutch insurance company expects a claim of 10% of the total sum insured when a building is flooded by 2 meters.

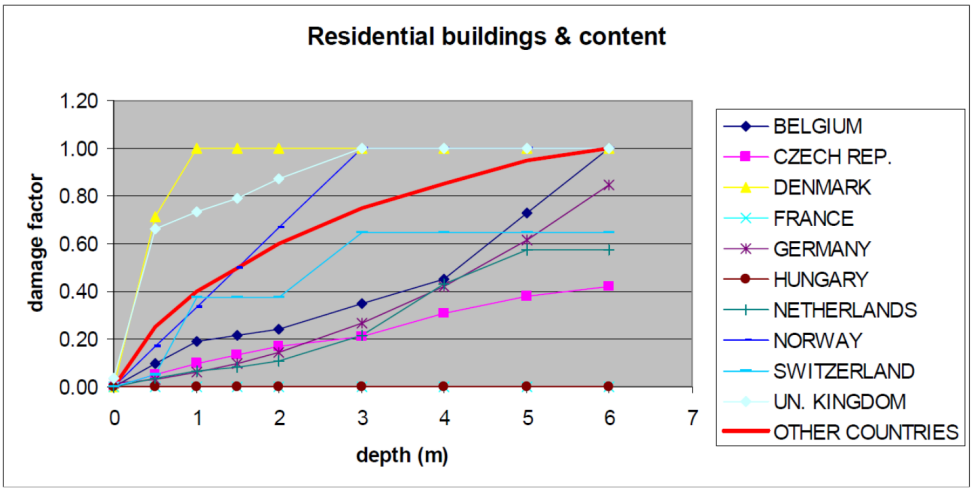


Figure 4: Depth-damage curves for a selection of European countries, gathered from Huizinga et al. (2017). Damages are considered as percentage of the total sum insured for residential buildings and include inventory damages.

For simplicity, we do not go into different damage categories and hence consider every building to be residential. We deem this assumption reasonable due to two reasons. One, residential buildings within the Netherlands take up to over 87% of total buildings (Statistics Netherlands, 2023a). Two, depth-damage curves show that maximum damage costs are lower for commerce and industry (Huizinga et al., 2017). This indicates that we do not neglect huge costs and at worst overestimate such damages slightly, which is a minor issue in the context of maximum flood losses. A last note is that the depth-damage curve is based on residential buildings and content. We are aware that this may differ from the WOZ value that we use in this thesis. However, the WOZ value serves as a good proxy and our methodology allows for insurance companies to plug in their own depth-damage curves easily into the model. Moreover, while we the WOZ value does not include content, it overestimates the total insured sum for the residential building without content as the total insured sum would only consider the rebuilding value.

4 Methodology

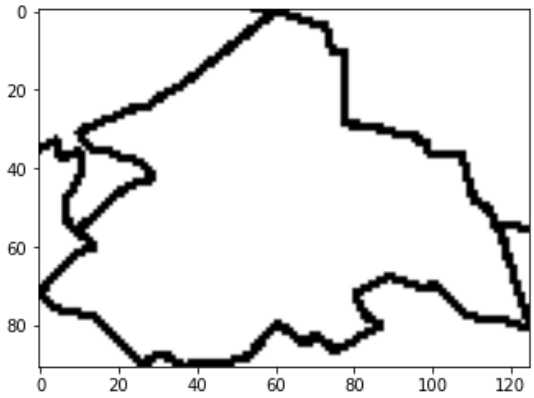
In this Section, we define notation and describe the methods used for answering the research questions. First, we extensively define the procedure of extracting damage proxies from the flood scenario maps. Secondly, we describe the four distributions and how we perform the quantile fitting analysis with different error metrics. Next, we precisely show the relevance of this research through an application. Finally, we show derivations for analytical optimization of the mean-variance hedge for insurers' perspective when issuing CAT bonds and shortly dive into some effects of CAT bond parameters for which no results are necessary.

4.1 Flood Scenarios to Damage Proxies

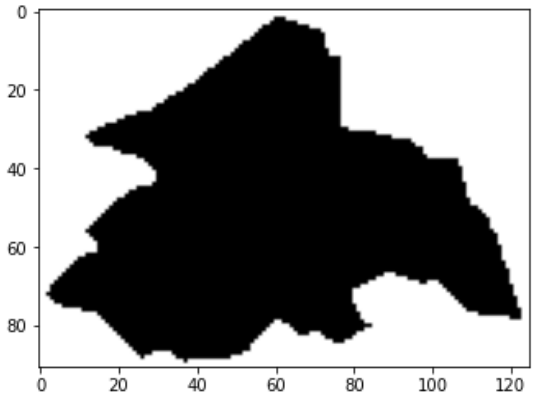
First, we translate the scenario maps to damage estimates. Due to time restrictions, we perform our analysis on South-Holland, a province of the Netherlands. This changes nothing to the procedure itself as the methodology is the same for every province within the Netherlands. The data preparation requires many assumptions and confirms the challenges due to data scarcity within the field of flood risk modeling. We keep track of every assumption that is necessary for translating the flood risk scenario maps to damages and what consequences follow. For simplification purposes, we only consider the data quantifying process for one municipality in this methodology: Gouda (remember Figure 3). Keep in mind that the data quantifying process is equal for all other municipalities within the Netherlands.

4.1.1 Image Recognition Algorithm

We use `matplotlib.image` in Python to convert images to pixels that store the pixel position and RGB-color codes. This is an essential first step for differentiating between different colors within the image and to see which fraction of buildings falls victim to a flood within different scenarios. We start by defining the borders of the municipality and use a self written flood-fill algorithm (like the bucket tool in paint) to determine all of the pixels that lie within the borders of the municipality. Figure 5 shows pixel data of the municipality Gouda.



(a) Data translation to pixel data for the municipality Gouda, where black pixels represent the borders of the municipality. The axes correspond to the pixel positions.

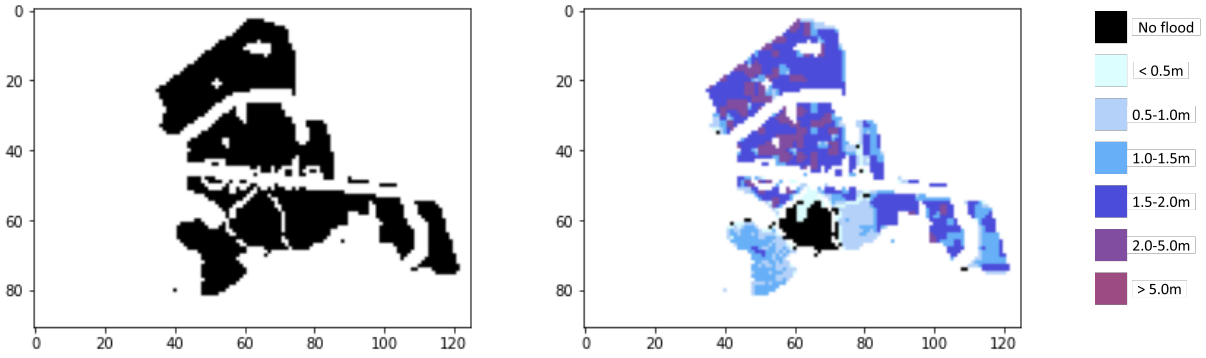


(b) Data translation to pixel data for the municipality Gouda, where black pixels represent the area within the borders of the municipality. The axes correspond to the pixel positions.

Figure 5: Representations of the borders and area within the borders of the municipality Gouda after running the flood-fill algorithm on the screenshots of the underlying map of the municipality.

After we know the area that lies within the borders of the municipality, we determine every pixel that corresponds to buildings - grey RGB color codes - within this area. The total number of pixels that corresponds to buildings is considered the full amount of buildings within the

municipality at hand. Gouda counts 38157 residential buildings in total. If for example 100 pixels correspond to buildings based on RGB color codes, one pixel corresponds to 381.57 buildings. After running the algorithm, the pixels corresponding to buildings show in Figure 6a. Lastly, we load the screenshots for all flood scenarios and divide every pixel in a certain depth range by color codes. Since we exclusively consider flood damage to buildings, we only determine flooded pixels for pixels that show black in 6a. One example, the 0.999 probability flood scenario for Gouda shows in Figure 6b.



(a) Pixel data for Gouda. Black pixels contain one or multiple buildings. The axes correspond to the pixel positions.

(b) 1 in 1000 year flood scenario pixel data for Gouda. Each color represents a maximum flood depth for each pixel that contains buildings.

Figure 6: Figure with representations of all the buildings in Gouda and (un)affected buildings in the 0.999 probability flood scenario.

Figure 6b assigns different maximum flood depth ranges to pixels that contain buildings and retrieves percentages of pixels at risk for different maximum flood depths.

4.1.2 Quantifying Pixel Data

Now, we estimate damage proxies from the obtained pixel data. We quantify the pixel data for every flood scenario separately by multiplying the pixel percentages per depth with their respective damage factors - percentages of the insured sum - of Huizinga et al. (2017). We consider the average damage factor within the maximum flood depth ranges. For example, for maximum flood depth $1.0 - 1.5$m, we consider the damage factor for 1.25m. Next, we multiply the sum of these fractions with the total amount of buildings and the average WOZ values for the municipality. This concludes the data quantifying process as we obtain the amount damages in euros for each flood scenario for a particular municipality. For all of South-Holland, we retrieve damage proxies by aggregating results for every individual municipality within South-Holland. Assessing flood risk for a province seems the most plausible, due to substantial differences in flood risk among provinces. This leads to differences in policies among provinces, where some

provinces need to consider multiple dike rings and rivers, while other provinces have almost no flood risk at all. This is the same for insurance companies that often focus on broader areas opposed to the whole country or just municipalities. However, the data quantifying process is flexible and can also be performed on a (multiple) multiplicity level or national level.

Additionally, we perform the calculation where we treat South-Holland as a municipality. This is possible since we know the average WOZ value for South-Holland. We perform this last calculation for comparison purposes and if results are relatively equal to aggregated results, this saves up time in the data preparation and we are able to perform analysis for the Netherlands as a whole. However, we hypothesize that results will differ due to the varying range of WOZ values per municipality (Table 1).

Lastly, we make notation more general for obtaining damage proxies for an arbitrary Dutch province by aggregating results for municipalities. Let Φ be the set of provinces in the Netherlands. Let $\phi \in \Phi$ be some province and denote flood scenario quantile $q \in (0.9, 0.99, 0.999, 0.9999)$. Given a set of municipalities in province ϕ: M_ϕ and a set of possible maximum flood depths D, the calculation procedure for damage proxies for a municipality and province follow in Equation (1) and (2) respectively. For an arbitrary municipality $m \in M_\phi$:

$$\hat{L}_{(m,q)} = \sum_{d \in D} x_{(m,d,q)} \cdot df_d \cdot b_m \cdot woz_m, \tag{1}$$

$$\hat{L}_{(\phi,q)} = \sum_{m \in M_\phi} \hat{L}_{(m,q)}, \tag{2}$$

with $\hat{L}_{(m,q)}$ and $\hat{L}_{(\phi,q)}$ the estimated yearly losses - damage proxies - for municipality m and province ϕ and scenario quantile q respectively. Moreover, where $x_{(m,d,q)}$ the fraction of pixels corresponding to buildings in municipality m and maximum flood depth d in scenario quantile q. Additionally, df_d denotes the (interpolated) damage factors of Huizinga et al. (2017) that correspond to maximum flood depth d. Lastly, b_m and woz_m represent the total buildings and average WOZ values of municipality m respectively. We perform an alternative estimation approach by using damages per m^2, which we elaborate upon in Appendix A. However, we choose to solely use the damage factor approach, since this includes more municipality specific information such as average WOZ values.

4.1.3 Application for Underwriting

For an underwriters' perspective, it is not only relevant to understand possible damages within a municipality or province, but rather potential damages within its client portfolio consisting of

residential buildings. We create two client portfolios of 10 addresses within South-Holland, P_1 and P_2. For privacy reasons, we do not mention the exact postal codes and housing numbers. However, we describe the methodology for obtaining the addresses so replication is possible nevertheless. We consult Funda (2023) for houses on sale filtered within the last 24 hours. We use this method to include randomness within the portfolios. Additionally, this method enables for using market values of the residential buildings instead of using the average WOZ value for the municipality. The use of market values allows for more specific and accurate results. Be aware of implied bias when using houses on sale at Funda, since this potentially excludes areas with low sales or with sales through alternative channels rather then through Funda.

We form portfolios P_1 and P_2 at 3pm September 9th 2023 and 12pm September 20th 2023 respectively. We then find the pixels that correspond to all ten addresses within the portfolios and retrieve four quantiles with corresponding losses for each portfolio. This provides insurers with clarity on their risk exposure. We also show quantile fitting and simulation for both portfolios. now use the program to simulate damage quantiles specifically for the portfolio of interest. This proves the versatility of the flood-to-damages procedure and shows that the procedure is applicable in multiple scenarios, both nationally and on smaller scales.

4.2 Damage Proxies to Loss Distribution

After obtaining estimates for damage proxies, we fit multiple distributions to the quantiles of the respective damage proxies to get a distribution for losses L. One could consider a variety of distributions to fit to the quantile data. we expect high losses far in the right tail, growing as the quantiles get closer to one. Hence, We opt for distributions with heavy tails to best imitate this tail behaviour. In this thesis, we consider four possible distributions for L, namely the classic Student t-distribution and a skewed Student t-distribution. Moreover, we consider the GPD and the GEV distribution. Define the set of distributions in the same order as above: $\mathcal{D} = (st, sst, gpd, gev)$ for notational purposes.

4.2.1 Distribution Notation

We first consider the classic Student t-distribution for which the probability density function (PDF) denoted as $f_{st}(\cdot)$ formulates as in Equation (3).

$$f_{st}(l; \mu, \sigma, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\sigma\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{1}{\nu}\left(\frac{l-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}, \tag{3}$$

with for now, losses $l \in \mathbb{R}$ and where location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma \in \mathbb{R}_{>0}$ and

degrees of freedom $\nu \in \mathbb{R}_{\geq 1}$[1]. When $\nu = \infty$, the distribution reduces to a Gaussian distribution with mean μ and variance σ^2.

Next, we follow the procedure of quantile fitting from Adrian et al. (2019), where they use with a skewed version of the classic univariate Student t-distribution, developed by Azzalini and Capitanio (2003). The PDF of the univariate skewed Student t-distribution formulates as in equation (4).

$$f_{sst}(l;\mu,\sigma,\alpha,\nu) = \frac{2}{\sigma}\, f_{st}\left(\left(\frac{l-\mu}{\sigma}\right);\nu\right)\; F_{st}\left(\alpha\left(\frac{l-\mu}{\sigma}\right)\sqrt{\frac{\nu+1}{\nu+\left(\frac{l-\mu}{\sigma}\right)^2}};\nu+1\right), \tag{4}$$

where shape parameter $\alpha \in \mathbb{R}$ and $F_{st}(\cdot)$ the cumulative distribution function (CDF) of the classic Student t-distribution. The special case $\alpha = 0$ reduces this distribution to the classic Student t-distribution. Mitchell et al. (2022) argue that the parametric assumption of using the univariate skewed Student t-distribution delivers gains compared to (mixed) Gaussian distributions when the true density is unimodal. Floods are irregular events and associated yearly damages should generally be close to zero. Therefore, we hypothesize unimodality around zero for the real loss distribution. This supports the idea of using a skewed Student t-distribution to fit the flood damage quantiles. The true density may also be bimodal in the case that losses are centered around the same number, introducing a second peak. However, we expect damages to vary per flooding and therefore unimodality of the true density.

Lastly, we consider two distributions that are often used in the context of heavy tailed (flood) data, namely the GPD and the GEV distribution (Wang, 1991; Morrison and Smith, 2002). The GEV distribution combines the Gumbel, Fréchet and Weibull distributions, which all classify as extreme value distributions. Properties of the GPD and GEV distribution are well established. Both distributions include a location, scale and shape parameter $\xi \in \mathbb{R}$. The PDF for the GPD shows in Equation (5).

$$f_{gpd}(l;\mu,\sigma,\xi) = \begin{cases} \frac{1}{\sigma}\left(1+\frac{\xi(l-\mu)}{\sigma}\right)^{-\frac{1}{\xi}-1}, & \text{if } \xi \neq 0 \\ \frac{1}{\sigma}\exp\left(-\frac{l-\mu}{\sigma}\right), & \text{if } \xi = 0 \end{cases}. \tag{5}$$

Additionally, the PDF for the GEV distribution formulates as follows:

[1] Hodges (2013) show that the degrees of freedom parameter is not strictly restricted to be integer.

$$f_{gev}(l;\mu,\sigma,\xi) = \begin{cases} \frac{1}{\sigma}\exp\left(-\left(1+\xi\frac{l-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right), & \text{if } \xi \neq 0 \\ \frac{1}{\sigma}\exp\left(-\exp\left(-\frac{l-\mu}{\sigma}\right)\right), & \text{if } \xi = 0 \end{cases}, \tag{6}$$

where the parameter space of loss l depends on ξ in the following way:

$$\begin{aligned} &l \in [\mu - \frac{\sigma}{\xi}, \infty) , \quad \text{if } \xi > 0 , \\ &l \in \mathbb{R} , \quad \text{if } \xi = 0 , \\ &l \in (-\infty, \mu - \frac{\sigma}{\xi}] , \quad \text{if } \xi < 0 . \end{aligned} \tag{7}$$

A last important general note is that damages cannot be negative. Therefore, the distributions should be truncated from the left at zero. This indicates the universal restriction $l \in \mathbb{R}_{\geq 0}$, of course also keeping in mind constraints on the range of l specific to the distributions. To obtain truncated versions of the four PDFs, we divide the PDF of each distribution $i \in \mathcal{D}$ by $1 - F_i(0;\theta)$, where $F_i(\cdot)$ denotes the CDF of distribution i with corresponding parameter set θ_i (Jamalizadeh et al., 2009). This division guarantees the PDFs to integrate to one.

To our best knowledge, `R` packages only offers the possibility of truncation only for the classic Student t-distribution (`crch`). Numerically optimizing truncated versions for the other distributions is extremely time consuming. For this reason, we only use the truncated version of a Student t-distribution. For the other distributions, we truncate the distributions where we use the parameter estimates for the non-truncated versions and assess whether the truncated versions remain adequate. Additionally, we investigate the consequences of using non-truncated distribution fits and setting negative values to zero when simulating.

Another possibility is to apply a log-transformation by for example considering the losses to be log-normally distributed. However, while parameters are clear for the log-normal distribution, such a well known distribution do not exist for heavy-tailed distributions like the GPD and GEV distributions. Moreover, it is not possible to first do a log-transformation on the data and then fit the heavy-tailed distribution, since the distribution fit to the log of the quantiles still needs to be truncated. Hence, in this thesis we continue with the truncation approach or the simulating and setting negative values to zero approach discussed above.

4.2.2 Distribution Moments

Information on the first and second moments for the four distributions are essential for the mean-variance hedge that follows later in this thesis. An inconvenient yet important note is that the

moments do not exist under specific circumstances. To prevent confusion hereafter regarding the location parameter and the first moment, we denote the first moment as $E[L] = \mu_i$ for $i \in \mathcal{D}$.

For the skewed and classic Student t-distribution, both the mean and variance are undefined if $\nu \leq 1$. Additionally, the variance is equal to ∞ if $1 < \nu \leq 2$. For the GPD, the mean is undefined if $\xi < 1$ and the variance is undefined for $\xi < \frac{1}{2}$. Finally, for the GEV distribution, it holds that $\mu_{gev} = \infty$ if $\xi \geq 1$ and the variance is equal to ∞ if $\xi \geq \frac{1}{2}$. Note that the existence of the moments depends only on ν and ξ. We keep this in mind when obtaining quantile fit distribution parameter estimates.

4.2.3 Quantile Fitting

Adrian et al. (2019) introduce quantile forecast estimates relating to the τ-th quantile of losses L as $\hat{Q}_\tau(L)$. We treat the quantile estimates retrieved from the flood maps as expert judgement. Hence, in our notation, the quantiles are not conditional on a regressor set. Additionally, their method considers quantiles at a finite number of τ, say $[\tau_1, \ldots, \tau_k]$, where $0 < \tau_1 < \tau_2 < \cdots < \tau_k < 1$ for which we use $k = 4$.

Adrian et al. (2019) and Mitchell et al. (2022) use the MSE for the quantile fitting procedure. Since flood risk quantiles might contain extreme values in the tail, this method might prove to be too sensitive to outliers. Additional other widely used error metrics are the MAE, MSPE and the MAPE (De Myttenaere et al., 2016; Armstrong and Collopy, 1992). The MAE is less prone to outlier data compared to the MSE due to the absence of squaring high values. The MSPE and MAPE measure the errors relative to their respective values and may therefore be a good choice as values may differ a lot over the quantiles. A drawback of the MSPE and MAPE is that values close to zero explode due to division, resulting in overvaluing such quantiles, which we keep in mind.

Therefore, we minimize the distance of the damage quantiles estimates with respect to the distribution quantiles of distribution i not by only using the MSE, but in four ways. We use the MSE, MAE, MSPE and MAPE as error metrics to compare the damage proxy quantile values and the theoretical quantile values. We examine outcomes based on the different scores and make a well-considered choice based on the results. The minimization formulas show in Equations (8) through (11).

$$\hat{\theta}_{\text{MSE},i} = \underset{\theta_i}{\arg\min} \sum_{\tau} \left(\hat{Q}_\tau(L) - F_i^{-1}(\tau;\theta) \right)^2, \tag{8}$$

$$\hat{\theta}_{\text{MAE},i} = \underset{\theta_i}{\arg\min} \sum_{\tau} \left| \hat{Q}_\tau(L) - F_i^{-1}(\tau;\theta) \right|, \tag{9}$$

$$\hat{\theta}_{\text{MSPE},i} = \underset{\theta_i}{\arg\min} \sum_{\tau} \left(\frac{\hat{Q}_\tau(L) - F_i^{-1}(\tau;\theta)}{\hat{Q}_\tau(L)} \right)^2, \tag{10}$$

$$\hat{\theta}_{\text{MAPE},i} = \underset{\theta_i}{\arg\min} \sum_{\tau} \left| \frac{\hat{Q}_\tau(L) - F_i^{-1}(\tau;\theta)}{\hat{Q}_\tau(L)} \right|, \tag{11}$$

where $F_i^{-1}(\cdot)$ denotes the inverted CDF of distribution $i \in \mathcal{D}$.

Due to the choice of $k = 4$, there is a restriction for fitting distributions. $|\theta| \leq 4$ is required since more parameters result in identification issues. In the case of the skewed Student t-distribution, we estimate exactly four parameters, which means an exactly identified cross sectional regression fit of the quantile estimates from flood maps on the quantiles of the skewed Student t-distribution. All other distributions have less than 4 parameters and thus do comply with the restriction due to the choice of k.

4.3 CAT Bonds

We now possess distribution fits for yearly losses L. Next, we examine CAT bonds as a reinsurance pool. We examine different parameters and payout structures and discuss their advantages and disadvantages to get insights on whether CAT bonds deem a viable option for the Netherlands when reinsuring flood risk for residential buildings.

We work with yearly losses due to the nature of the flood scenario maps. We assume yearly losses to be identically and independently distributed from the distribution fits. CAT bonds may have maturities of more than one year. Hence, we define the agregated loss process for a time-period of T years, that formulates as in Equation (12).

$$L^{(T)} = \sum_{t=1}^{T} L_t, \tag{12}$$

where L_t denote the losses for year t. Observe that $L^{(1)} = L_1$.

The pricing structure of CAT bonds is fundamentally a function of (aggregated) losses $f(L^{(T)})$, where we often consider expected losses $E[L^{(T)}]$ (Zonggang and Ma, 2013). CAT bond prices typically associate with an attachment point a. This point is a certain threshold for which policy holders become eligible to receive compensation. When losses increase towards the attachment point, the probability of investors not receiving their full payment back increases.

Consequently, this is why expected losses pay a role in the pricing structure of CAT bonds. We set the range for the attachment point $a \in \mathbb{R}_{\geq 0}$, since losses are nonnegative. These fundamental insights are also relevant for insurance companies. In this thesis, we focus on the insurers' perspective of CAT bond payoffs and effects of changing parameters within the bond structure. Note that really small values of a are not logical for the bonds, unless an insurance company wants to completely eliminate their risk irrespective of the costs of issuing bonds.

We follow notation of Zonggang and Ma (2013). Consider a zero-coupon CAT bond, with principal y in euros, maturity T and attachment point a. Define P_{CAT}, a random variable that corresponds to the payout fraction of the CAT bond that is paid to the investor. Observe that this variable is independent of the CAT bond principal or premium paid by the insurance company. Mathematical notation shows in Equation (13).

$$P_{\text{CAT}} = \begin{cases} p & ; \quad L^{(T)} > a \\ 1 & ; \quad L^{(T)} \leq a \end{cases}, \tag{13}$$

where $p \in [0, 1)$ represents the portion of the principal that investors receive when losses surpass the attachment point a. Additionally, the insurance company pays a premium r to the investor in return for taking over a piece of the risks, depending on the attachment point. An alternative representation of P_{CAT} shows in Equation (14).

$$P_{\text{CAT}} = p\mathbb{I}(L^{(T)} > a) + \mathbb{I}(L^{(T)} \leq a). \tag{14}$$

$\mathbb{I}(\cdot)$ an indicator function that takes either value one or zero depending on whether the event in the brackets occurs.

Now consider a random variable that denotes the payoff for the insurers' perspective. We exclude collected premiums that clients pay in return for insurers, since this is some fixed constant this is irrelevant for the maximization and determination for the amount of CAT bonds to be issued. We define insurers' payoff P_{INS} as follows:

$$P_{\text{INS}} = -L^{(T)} + \beta y - \beta y(1 + r)P_{\text{CAT}}, \tag{15}$$

where $\beta \in \mathbb{Z}_{\geq 0}$ denotes the amount of issued CAT bonds, P_{CAT} as defined above and with y the price for one CAT bond in billions of euros. For example, $y = 1$ corresponds with a CAT bond being worth one billion euros. The positive term that includes β denotes the money received through principals paid by the investors, while the negative term that includes β denotes the money paid back to investors, depending on P_{CAT}. This provides for a quantifying way to

determine whether CAT bonds are a reasonable choice for insurance companies to reinsure flood risk, since $\beta = 0$ implies that the optimal choice for the insurer is to not issue any CAT bonds.

4.3.1 Mean-Variance Hedging

We use mean-variance hedging, originally proposed by Markowits (1952), to determine an optimal strategy for insurance companies regarding the issuing of CAT bonds. The objective is to maximize the payoff for the insurer P_{INS}, while keeping risks to a minimum. For this maximization problem (Schweizer, 2010), we maximize the insurers' payoff for β:

$$\max_{\beta} E[P_{\text{INS}}] - \gamma \text{Var}(P_{\text{INS}}), \tag{16}$$

where γ denotes the risk aversion of the insurer, with higher values for γ indicate a higher risk aversion of the insurance company. Intuitively, the expected payoff is negative since we exclude client premiums from the equation. observe that increasing γ has a negative impact on the maximization function as the variance is always non-negative.

In reality, insurance companies are restricted to decide on issuing CAT bonds at current time and cannot adjust their initial strategy over the years. They may choose to issue additional CAT bonds in following years, when next year is now current time. However, this excludes dynamic decision making, which allows for relatively simpler maximization.

4.3.2 Analytical Derivations

To analytically solve the mean-variance hedge, we require derivations for both moments in the maximization function from Equation (16). We provide analytical derivations for maturity $T = 1$. For notational convenience, denote $L = L^{(1)}$, with mean $E[L] = \mu_i$, variance $\text{Var}(L) = \sigma_i^2$ and CDF $F_i(l)$, where $i \in \mathcal{D}$. We make the following derivations for the moments of interest with maturity $T = 1$:

$$\begin{aligned} E[P_{\text{INS}}] &= -E[L] + \beta y - \beta y(1+r)E[P_{\text{CAT}}] \\ &= -\mu_i + \beta y - \beta y(1+r)(P(L \leq a) + pP(L > a)) \\ &= -\mu_i + \beta y - \beta y(1+r)(F_i(a) + p(1 - F_i(a)), \end{aligned} \tag{17}$$

$$\text{Var}(P_{\text{INS}}) = \text{Var}(L) + \beta^2 y^2 (1+r)^2 \text{Var}(P_{\text{CAT}}) + 2\beta y(1+r)\text{Cov}(L, P_{\text{CAT}}). \tag{18}$$

Since L and P_{CAT} are not independent, the variance expression from Equation (18) requires some more work. Let us first derive the expression for $\text{Var}(P_{\text{CAT}})$.

$$\begin{aligned}
\text{Var}(P_{\text{CAT}}) &= E[\text{Var}(P_{\text{CAT}}|L)] + \text{Var}(E[P_{\text{CAT}}|L]) \\
&= \text{Var}(E[P_{\text{CAT}}|L]) \\
&= E[E[P_{\text{CAT}}|L]^2] - E[E[P_{\text{CAT}}|L]]^2 \\
&= p^2(1 - F_i(a)) + F_i(a) - (p(1 - F_i(a)) + F_i(a))^2 \\
&= -F_i(a)^2p^2 + 2F_i(a)^2p + F_i(a)p^2 - F_i(a)^2 - 2F_i(a)p + F_i(a),
\end{aligned} \tag{19}$$

where we use the fact that $E[\text{Var}(P_{\text{CAT}}|L)] = 0$ as P_{CAT} is a fixed number if L is known.

An alternative approach to get an expression for $\text{Var}(P_{\text{CAT}})$ follows from using the alternative representation from Equation (14). The indicator functions are Bernoulli distributed random variables $\mathbb{I}(L \leq a) \sim B(F_i(a))$ and $\mathbb{I}(L > a) \sim B(1 - F_i(a)) = 1 - \mathbb{I}(L \leq a)$. We then derive the variance of the CAT bond payoff for the insurer as follows.

$$\begin{aligned}
\text{Var}(P_{\text{CAT}}) &= \text{Var}(p\mathbb{I}(L > a) + \mathbb{I}(L \leq a)) \\
&= p^2\text{Var}(\mathbb{I}(L > a)) + \text{Var}(\mathbb{I}(L \leq a)) + 2p\text{Cov}(\mathbb{I}(L > a), \mathbb{I}(L \leq a))) \\
&= p^2F_i(a)(1 - F_i(a)) + F(a)(1 - F_i(a)) \\
&\quad + 2p(E[\mathbb{I}(L \leq a)\mathbb{I}(L > a)] - F_i(a)(1 - F_i(a))) \\
&= p^2F(a)(1 - F_i(a)) + F(a)(1 - F_i(a)) \\
&\quad + 2p(E[\mathbb{I}(L \leq a)(1 - \mathbb{I}(L \leq a)] - F_i(a)(1 - F_i(a))) \\
&= p^2F_i(a)(1 - F_i(a)) + F_i(a)(1 - F_i(a)) - 2p(F_i(a)(1 - F_i(a)),
\end{aligned} \tag{20}$$

where we make use of the properties of the Bernoulli distribution and the fact that the second moment is equal to it's first moment. The last expression simplifies exactly to the result in Equation (19).

Next, we derive the covariance in Equation (18). The derivation shows in Equation (21).

$$\begin{aligned}
\text{Cov}(L, P_{\text{CAT}}) &\overset{(14)}{=} \text{Cov}(L, p\mathbb{I}(L > a) + \mathbb{I}(L \leq a)) \\
&= E[Lp\mathbb{I}(L > a) + L\mathbb{I}(L \leq a))] - \mu_i(p(1 - F_i(a)) + F(a)) \\
&= \int_0^a l f_i(l)dl + p\int_a^\infty l f_i(l)dl - \mu_i(p(1 - F_i(a)) + F(a)),
\end{aligned} \tag{21}$$

where the first integral starts at 0 due to distribution truncation. For non-truncated distributions, this first integral starts at $-\infty$. Note that the integrals and μ_i depend on the particular distribution functions of distribution $i \in \mathcal{D}$. We expect $\text{Cov}(L, P_{\text{CAT}})$ to be negative, since higher losses increase $\text{P}(L > a)$, which decreases the value of P_{CAT} since $p < 1$.

We now have the full derivation of Equation (18). Aside from this derivation, we now optimize the resulting minimization equation (24). We do this derivation analytically by taking the derivative of the objective function with respect to β and setting the derivative equal to zero. First work out the function

$$\begin{aligned} f(\beta) &= E[P_{\text{INS}}] - \gamma \text{Var}(P_{\text{INS}}) \\ &= -\mu_i + \beta y - \beta y(1+r)(F_i(a) + p(1 - F_i(a))) \\ &\quad - \gamma(\text{Var}(L) + \beta^2 y^2 (1+r)^2 \text{Var}(P_{\text{CAT}}) + 2y(1+r)\beta \text{Cov}(L, P_{\text{CAT}}), \end{aligned} \tag{22}$$

where we take the derivative and set it equal to zero as follows.

$$\begin{aligned} f(\beta)' = y - y(1+r)(F_i(a) + p(1 - F_i(a))) - 2\gamma\beta y^2 (1+r)^2 \text{Var}(P_{\text{CAT}}) \\ - 2\gamma y(1+r)\text{Cov}(L, P_{\text{CAT}}) = 0, \end{aligned} \tag{23}$$

which we solve for β to find the optimal amount of CAT bonds to be issued from the insurer's perspective. We denote the optimal β as

$$\beta^* = \max\left\{0, \frac{\frac{1}{(1+r)} - (F_i(a) + p(1 - F_i(a))) - 2\gamma \text{Cov}(L, P_{\text{CAT}})}{2y(1+r)\gamma \text{Var}(P_{\text{CAT}})}\right\}, \tag{24}$$

where β^* is the optimal solution of the mean-variance hedge and hence the optimal number of CAT bonds with maturity $T = 1$ to issue for the insurer. Hereafter, we refer to $\frac{1}{(1+r)} - (F_i(a) + p(1 - F_i(a)))$ as the expectation term, $-2\gamma \text{Cov}(L, P_{\text{CAT}})$ as the covariance term and $2y(1+r)\gamma \text{Var}(P_{\text{CAT}})$ as the variance term. Note that we make sure that $\beta^* \geq 0$ by taking the maximum of the expression and zero.

An important last note is that, as stated in Section 4.2.2, $E[L]$ and $\text{Var}(L)$ may not exist under some circumstances. At first glance, these terms are not present in the calculation for β^*. However, a nonexistent mean results in divergent or non-solvable integrals in the calculation of $\text{Cov}(L, P_{\text{CAT}})$. Hence, a nonexistent mean results in no possible analytical solution for the maximization problem. We may obtain β^* via simulation results of the distribution of L. Note

that this problem with the integrals relates to the $E[L]$ and that a nonexistent $\text{Var}(L)$ does not evoke any problems in the calculation of β^*, as this term does not appear in Equation (24).

4.3.3 Effects of CAT bond parameters

Consider the derivations from the previous Section. On forehand before obtaining results, we explore causal effects on β^* when changing individual parameters, ceteris paribus. Parameters to consider are y, γ, r, p and a. Keep in mind that the maturity is set to one year.

Starting with y, this parameter shows only once in Equation (24), which is in the denominator. Therefore, increasing y (the bond price) decreases the amount of bonds to be issued by the same factor. This makes sense as the bond price does not matter as a 10 times higher price would result in 10 times less issued bonds and results to the same optimum. Since β can only be integer, lower bond prices provide more accurate results since there is less chance of rounding β^* after optimization.

Next, consider the risk aversion parameter γ. This parameter appears twice in the expression for β^*. Using Equation (24), we obtain the following limits for γ:

$$\begin{aligned} \lim_{\gamma \to \infty} \beta^* &= \max\left\{0, \frac{-\text{Cov}(L, P_{\text{CAT}})}{y(1+r)\text{Var}(P_{\text{CAT}})}\right\}, \\ \lim_{\gamma \to 0^+} \beta^* &= \max\left\{0, \pm\infty\right\}, \end{aligned} \tag{25}$$

where the second limit is either positive infinity when the expectation term is positive and zero if the expectation term is negative. Observe that when the risk aversion parameter goes towards infinity, the optimal decision in the limit only depends on the variance and covariance terms, neglecting the expectation term. Contrarily, when γ approaches zero, the optimal decision depends solely on whether the fixed part is profitable (issue infinite bonds) or not profitable (issue no bonds), which depends on the premium and attachment point. An additional note is that β^* cannot be negative as an insurer cannot issue a negative amount of CAT bonds, so the limit of β^* when γ approaches zero is either positive infinity or zero.

We now provide an important insight on the role of the risk aversion parameter γ in this particular mean-variance hedge. First, observe that for the first limit, β^* converges to a fraction with both the covariance and variance term. We expect $\text{Cov}(L, P_{\text{CAT}})$ to be negative, so the fraction becomes a positive number. Moreover, we established that β^* goes towards positive infinity or zero when γ goes to zero. Generally, we expect this to be zero opposed to positive infinity, since investors should receive a higher premium than the expected payoff due to uncertainties, such that $\frac{1}{(1+r)} - (F_i(a) + p(1 - F_i(a))) < 0$. Therefore, increasing γ leads to an optimal solution

of more CAT bonds to issue, converging to the first limit. This result makes sense intuitively, since a higher risk aversion for insurance companies implies that they wish to reallocate more risk towards investors via CAT bonds.

Now consider the premium r, the risk premium paid to investors for taking over a portion of the risk. We see two occurrences in the expression for β^*. Since $(1+r) \geq 1$ and using the fact that r only shows in denominators within the equation, a higher r results in a lower β^*. This makes sense as bonds are more costly to issue and hence less attractive for insurers when r increases.

The effects of a and p are less prevalent, since they appear in both the variance and covariance terms as well as the expression itself for β^*. Intuitively, we hypothesize that a higher payout fraction p in the case of a catastrophe leads to a lower β^*. Note that a higher p becomes closer to one, which decreases $\text{Var}(P_{\text{CAT}})$, which in turn would decrease β^*. This holds true in Equation (20), since $p^2 < 2p$ for every value of p. However, a higher payout fraction does not necessarily decrease β^*, since the result is also dependent on the relation between the expectation and covariance term in the nominator.

A higher attachment point corresponds to more risk exposure for insurance companies since the catastrophic event is defined for larger losses. For attachment point a, such that $F_i(a) = 0.5$, $\text{Var}(P_{\text{CAT}})$ is highest. Increasing or decreasing a from that point decreases the variance term towards zero as $F_i(a) \to 1$ or $F_i(a) \to 0$. Moreover, the covariance term also goes towards zero as a increases as it follows for $\text{Cov}(L, P_{\text{CAT}})$ that

$$\begin{aligned} \lim_{a \to \infty} \text{Cov}(L, P_{\text{CAT}}) &= \int_0^\infty l f_i(l) dl + p \int_\infty^\infty l f_i(l) dl - \mu_i = \mu_i + 0 - \mu_i = 0, \\ \lim_{a \to 0^+} \text{Cov}(L, P_{\text{CAT}}) &= \int_0^0 l f_i(l) dl + p \int_0^\infty l f_i(l) dl - \mu_i = 0 + p\mu_i - p\mu_i = 0, \end{aligned} \tag{26}$$

This together with the effects described for the variance term indicate that if a either decreases towards zero or increases towards positive infinity, β^* converges to the expectation term, blown up by the variance term. Due to β^* restricted to be nonnegative, this means that β^* converges to zero when the expectation term is zero or negative and converges to positive infinity when the expectation term is positive. The limits for the expectation are as follows:

$$\begin{aligned}\lim_{a\to\infty}\left(\frac{1}{(1+r)}-(F_i(a)+p(1-F_i(a)))\right)&=\frac{1}{(1+r)}-1,\\ \lim_{a\to 0^+}\left(\frac{1}{(1+r)}-(F_i(a)+p(1-F_i(a)))\right)&=\frac{1}{(1+r)}-p,\end{aligned}\tag{27}$$

where the first limit is generally negative since $r \geq 0$ and we expect the second limit to be positive due to r and p usually relatively small. Therefore, β^* is expected to become positive infinity for really small a and zero for large values of a that correspond to $F_i(a) \to 1$.

The effects regarding changes in the attachment point intuitively make sense as high values of a indicate essentially no variance or covariance since losses have almost no chance of exceeding the attachment point, making P_{CAT} close to a fixed number. A similar explanation follows for a close to zero, indicating that P_{CAT} is almost always equal to p.

4.3.4 Optimization by Simulation

Lastly, we perform the mean-variance hedge by simulation of losses. We perform the analysis for the chosen most appropriate distribution fit. Numerical optimization allows us to consider other maturities. Hence, we consider maturities ranging from 1 to 10. For optimization, we use the `optimize` function from `R` and set the seed to zero. This creates a small bias in simulation outcomes. However, this allows for replication. Moreover, the bias becomes negligible as the amount of simulations becomes large. For every maturity, we draw $1,000,000$ independent aggregated losses from which we estimate $\hat{\beta}^*$.

5 Results

In this Section, we provide the results after using the data and methodologies from previous sections. First, we show the translation of flood risk quantiles towards damages in euros. Next, we show a wide range of distribution quantile fits on the damage quantiles and make a selection of distribution fits which we further examine. Moreover, we show an alternative application on a more local level for insurers by using portfolios of residential buildings instead of looking at municipalities or provinces as a whole.

Furthermore, we show how issuing CAT bonds serves as a potential solution to partially reallocate risk for the insurer towards other parties and explore effects of changing variables that exist in such bonds. Finally, we show results for a sensitivity analysis regarding the zoom, performed on the quantifying process of the flood risk quantiles to damages.

5.1 Flood Scenarios to Damage Proxies

First, we translate Flood scenario maps to damage proxies by running the image recognition algorithm for all municipalities in South-Holland separately. We retrieve pixel data for all four flood scenario quantiles from the algorithm and estimate damage proxies following the methodology from Section 4.1.

5.1.1 Quantifying Pixel Data

All explicit results for every quantile for the individual municipalities and the aggregated results show in Appendix B. Visualization of the average damages for every municipality shows in Figure 7. Observe that the spread of values of individual municipalities grows larger as the quantiles go further into the right tail, most likely due to several municipalities being inside of the of flood range of dike rings collapsing, while other municipalities are never impacted by such floods. Moreover, the scale of municipalities plays a big role. For example, a big municipality Rotterdam covers considerably more buildings than smaller municipalities, which results in more potential damages.

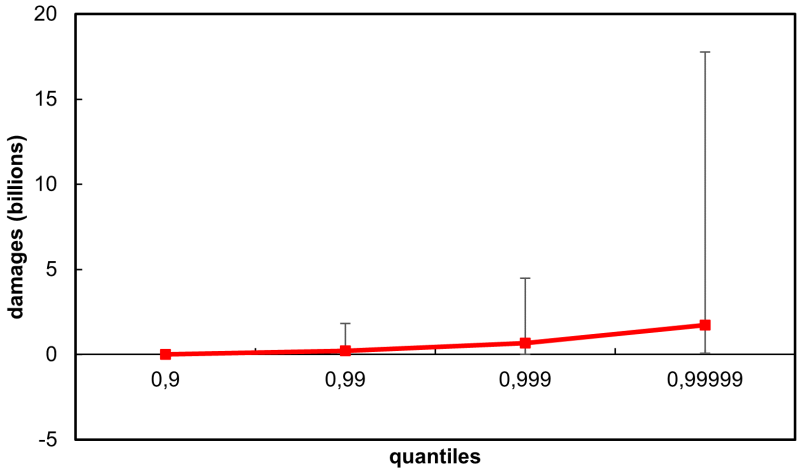


Figure 7: Average damages in billions for every municipality within South-Holland. The graph also shows the maximum and minimum values vertically. Note that the horizontal axis is not really linear in steps for visualization purposes.

We aggregate the damages calculated for the quantiles for all individual municipalities to obtain the total damages for every quantile for all of South-Holland. Define these total damages as the aggregated version for South-Holland. Additionally, we complete a similar calculation by using an alternative, averaging way, where we convert all of South-Holland to damages as a whole. This second calculation uses the average WOZ values for the province instead of for every municipality individually, which saves calculation time for the program. If both calculations

are relatively equal, running the program for provinces directly without aggregating individual municipalities would potentially be a good option for other provinces, assuming that the process for South-Holland is representative for other provinces. This speeds up the process substantially and allows us to examine other provinces too within the scope of this thesis. Results for the individual municipalities and the aggregated results show in Figure 8. Although the patterns between quantiles are fairly the same for both calculation procedures, the results itself clearly differ. This shows that performing calculations for municipalities individually and then aggregate results add value opposed to averaging for the province as a whole. For this reason, this thesis exclusively considers results for South-Holland. Note that results for the Netherlands as a whole follow the same procedure by aggregating all municipalities. A possible explanation for why the aggregated and averaged calculations vary is that some municipalities with high WOZ values (e.g. Wassenaar) are relatively safe from flood risk and therefore are less prone to damages to residential buildings, while the average WOZ value for the municipality drives up the average WOZ value for the province nevertheless.

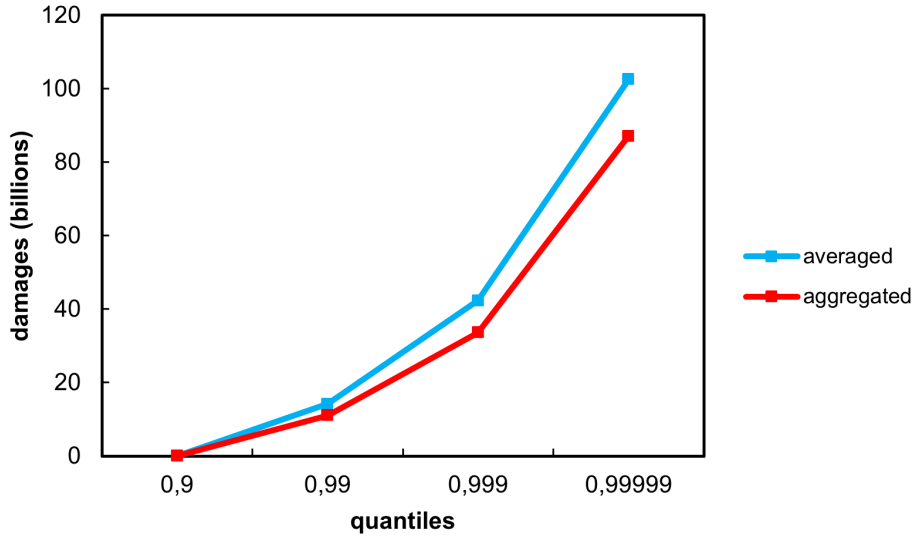


Figure 8: Aggregated flood damages in billions for different quantiles for every municipality within South-Holland. Flood damages show also averaged for the whole province in terms of WOZ values. Note that the horizontal axis is not really linear in steps for visualization purposes.

Damages differ substantially over the quantiles, ranging from 4 million euros to 87 billion euros. This shows the impact of floods due to dike breaches versus fluvial floods. To put the damages into perspective, consider the flood damages in Limburg, another province of the Netherlands in July 2021. All damages taken into account is estimated for a total of around 1.8 billion euros (NOS, 2022). This comparison is not entirely reasonable, since South-Holland is more densely populated. Moreover, this includes all damages and does not exclusively consider

damage to buildings. When damages to buildings are known, one could run the algorithm in this thesis for Limburg and estimate the corresponding flood probability. Nevertheless, this provides some indication for the order of magnitude of the damages.

5.1.2 Application for Underwriting

In this application, we consider a portfolio of residential buildings within South-Holland and take a look at the underwriters' perspective. This provides a more precise look opposed to the previous more generalized view on the national level regarding all residential buildings within a province. Moreover, this shows the relevance of this thesis and how the theory translates to real-world practicalities. We consult Funda to obtain portfolios P_1 and P_2 as discussed in Section 4.1.3. We use a similar approach as before to quantify flood risk to actual damages. Visualization for the distribution of residential buildings for portfolios P_1 and P_2 show in Figure 9. We extract data from the specific pixels of the locations and use a similar calculation as Equation (1) to estimate losses for each quantile. Remember that we use market values from Funda in contrast to average WOZ values of the municipalities to obtain more precise results.

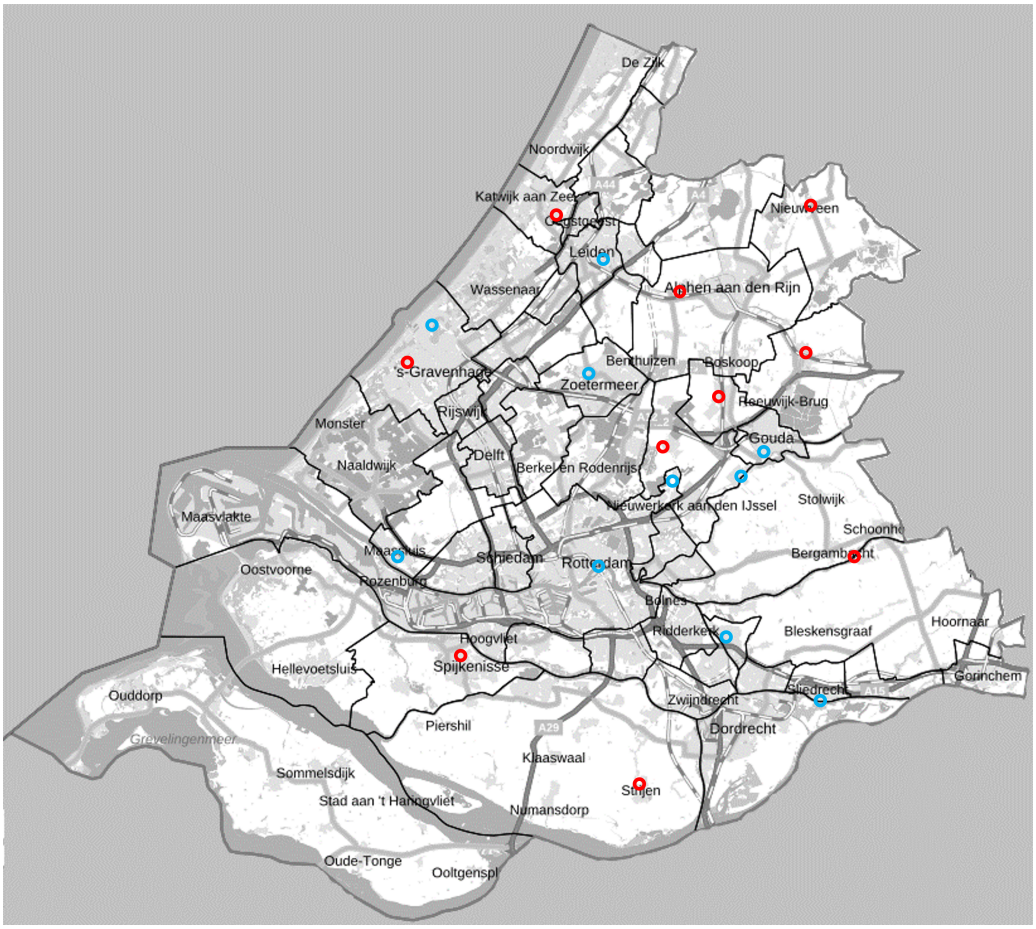


Figure 9: Visualization of the portfolio distribution that consists of 10 residential buildings within South-Holland. Locations of the residential buildings show in red and blue for P_1 and P_2 respectively.

Table 9 shows results for obtaining damage proxies for portfolios P_1 and P_2. This immediately shows the differences that may arise within specific portfolios. Since we only consider ten residential buildings per portfolio and only receive four quantiles, information is relatively fragile to perform quantile fitting with distributions. However, insurers may consider quantile fitting for larger portfolios. For such larger portfolios, insurers are able to calculate other measures such as expected shortfall after fitting a distribution. Even for smaller portfolios like we show, insurance companies are able to extract information from quantile results. Quantiles are equivalent to Value at Risk (VaR), a risk measure often used in the field of risk management. Insurers may either voluntarily or mandatorily use this risk measure due to regulation laws. For example, requirements for Insurers may include a safety net corresponding to a VaR(0.999). For P_1, this implies that the insurance company must have approximately 700 thousand euros readily available at all times.

Table 2: Summary statistics of the total amount of buildings and WOZ values for each municipality within the Netherlands

Portfolio	q0.9 (€)	q0.99 (€)	q0.999 (€)	q0.9999 (€)	Market Value (€)
P_1	0	6,638	346,633	707,246	4,596,000
P_2	0	61,845	260,870	523,676	5,949,000

Estimating damage proxies provides for a better awareness for insurers to access the risk within their portfolio and allows them to react accordingly. This goes beyond only regulatory capital restrictions. For example, an insurance company may use the same approach to assess the risk of adding new residential buildings to their portfolio. While insurers are prohibited to discriminate by declining new clients - residential buildings - to their portfolio, it provides the insurer with clarity on how risk exposure of the portfolio changes for the different quantiles.

Lastly, the damage proxies provide insights for insurers regarding decision making for their premiums paid by their clients. Insurers operate in a competitive environment against other insurance companies. Clarity on potential losses and probabilities allows insurers to make a better assessment of the expected claims. In combination with clarity on all other relevant business aspects such as underwriting expenses, operation costs and regulatory obligations, the insurer then estimates the premiums needed to at least cover all expenses.

5.2 Damage Proxies to Loss Distribution

After obtaining estimated losses in billions of euros corresponding to the four flood scenario maps, we perform quantile fitting for defined distribution set $\mathcal{D}$ to get distribution fits for the losses. Hereafter, we consider the aggregated calculation for damage losses in South-Holland to be the actual damage quantiles. We minimize with respect to the distance between the actual quantiles and the distribution quantiles. We perform this minimization for the four distributions based on multiple error metrics, namely the MSE, MAE, MSPE and MAPE.

5.2.1 Quantile Fitting

There are two options to obtain a distribution for the losses. The first method is to aggregate the damage quantiles and fit a distribution to the quantiles. The second option is to fit a distribution for every municipality individually and then to aggregate the distributions. An advantage of the second method is that it allows for the distribution to capture individual characteristics of multiplicities into the distribution. However, we see in the results that the quantiles vary strongly among multiplicities, with some municipalities having no damages until the worst case scenario

(0.9999 quantile). This results into very different distribution fits or different multiplicities may even require different theoretical distributions. Consequently, it becomes a difficult choice to give weights to the distribution parameters when aggregating. Furthermore, we only work with four quantiles, which means that the uncertainty of the distribution fits adds up even more when considering every municipality individually. Lastly, it is hard to model the dependencies between municipalities when aggregating the distributions, while this is more convenient when summing up the damage proxies. Hence, we first aggregate the damage proxies and then proceed with the quantile fitting using the aggregated proxies.

The quantile fitting results show in Table 3. Note that scores are only comparable relative to the same error metric. The best (lowest) scores within the same error metrics are highlighted in bold. The GEV distribution fits outperform the other distributions when using the MSE, MAE and MSPE. In terms of the MAPE, the skewed Student t-distribution scores best.

Table 3: Optimized loss distribution fits. The scores show the squared or absolute differences between the distribution fit quantiles and the actual proxy quantile values. The second row shows the actual damage proxy values for each quantile.

	Distribution	Score	Q0.9 (bn€)	Q0.99 (bn€)	Q0.999 (bn€)	Q0.9999 (bn€)
			0.0445	**11.1116**	**33.6455**	**87.1544**
MSE	SST	14.77	3.56	11.64	32.22	87.47
	ST	17.63	3.87	11.59	32.01	87.52
	GPD	11.05	3.19	11.41	32.63	87.38
	GEV	**10.36**	3.09	11.42	32.67	87.37
MAE	SST	10.21	2.67	9.04	28.18	87.2
	ST	10.01	2.82	9.2	28.36	87.19
	GPD	7.08	2.86	9.56	30.25	95.56
	GEV	**5.92**	3.59	12.66	34.43	87.11
MSPE	SST	1.76	0.0585	0.6043	6.0452	60.4521
	ST	1.78	0.0585	0.5899	5.8997	58.9974
	GPD	1.49	0.0486	0.6445	7.9585	97.7217
	GEV	**1.48**	0.0486	0.6697	8.14	97.9345
MAPE	SST	**2.96**	0.0445	0.1439	0.4256	1.2469
	ST	2.99	0.0445	0.0445	0.0445	0.0445
	GPD	2.99	0.0445	0.053	0.0547	0.055
	GEV	2.99	0.0445	0.0569	0.0601	0.061

Notes: Scores can only be compared for the same error metric. For the Score column, the best score shows in bold for each error metric.

In line with our hypothesis, the distribution fits that are based on the MSE and MAE assign a substantial portion of the weights towards the highest quantile value, while disregarding the low quantile values. Both the MSE and MAE distribution fit results in Table 3 estimate the 0.9 quantile value around 3 for every distribution, while the actual quantile value 0.0445 is smaller by

a factor of approximately 67. On the contrary, the MSPE and MAPE assign weights to quantiles relative to their respective values. A drawback is that values close to zero tend to get high weights due to division within the error metric by the quantile value itself. This especially shows in the MAPE results, where the fitted quantiles assign nearly all the weight to the first quantile value, where now the high quantiles are neglected instead. The most balanced error metric for the minimization seems to be the MSPE as it spreads the weights most towards the different quantile values. Although the second quantile is off by a factor 10, all quantiles are relatively well estimated and both the smallest and highest quantile values fit acceptably. For now, the MSE deems to be the best overall choice as an error metric for this particular application of quantile fitting. However, the MSPE based fits better satisfies the actual value for the 0.9. Therefore, we consider results for both the MSE and MSPE error metric hereafter.

The parameter estimates for the distribution fits based on error metrics MSE and MSPE show in Table 4. Observe the differences in σ, ξ and ν for both minimization procedures. A rather inconvenient observation is that the MSPE fitted parameter estimates indicate for all of the distributions that there does not exist a finite mean. This eliminates the possibility of analytically solving a mean-variance hedge since that requires the integrals in Equation (21) to be finite. The MSE fitted parameters do allow for analytically solving the mean-variance hedge, as the mean is finite for all distributions.

Table 4: Parameters of the optimized distribution fits for the MSE and MSPE error metrics.

	Distribution	μ	σ	ξ	ν
MSE	SST	0	2	0	2.31
	ST	0	1.44	-	2.29
	GPD	0	0.83	0.412	-
	GEV	0	0.835	0.411	-
MSPE	SST	0	0.019	0	1
	ST	0	0.00927	-	1
	GPD	0	0.00469	1.08894	-
	GEV	0	0.00507	1.07987	-

The GEV distribution fits perform best for both the MSE and MSPE based optimization procedures. Hence, we consider the GEV distribution and its parameter estimates hereafter. The GEV distribution PDFs for both error metrics show in Figure 10. We exclude the vertical axis since only the pattern for the PDFs are relevant here. The difference between the fitted distributions is inherently clear as the MSPE fit drops towards zero more swiftly than the MSE fit. Since the first actual quantile value is 0.0445 billion euros, additional more zoomed in graphs

also show for both distribution fits in Figures 10c and 10d. The MSPE based distribution fit is already in the tail within the area around this value, while the MSE based distribution fit is not even close to the tail for this value: $F_{gev,MSE}(0.0455) = 0.031$.

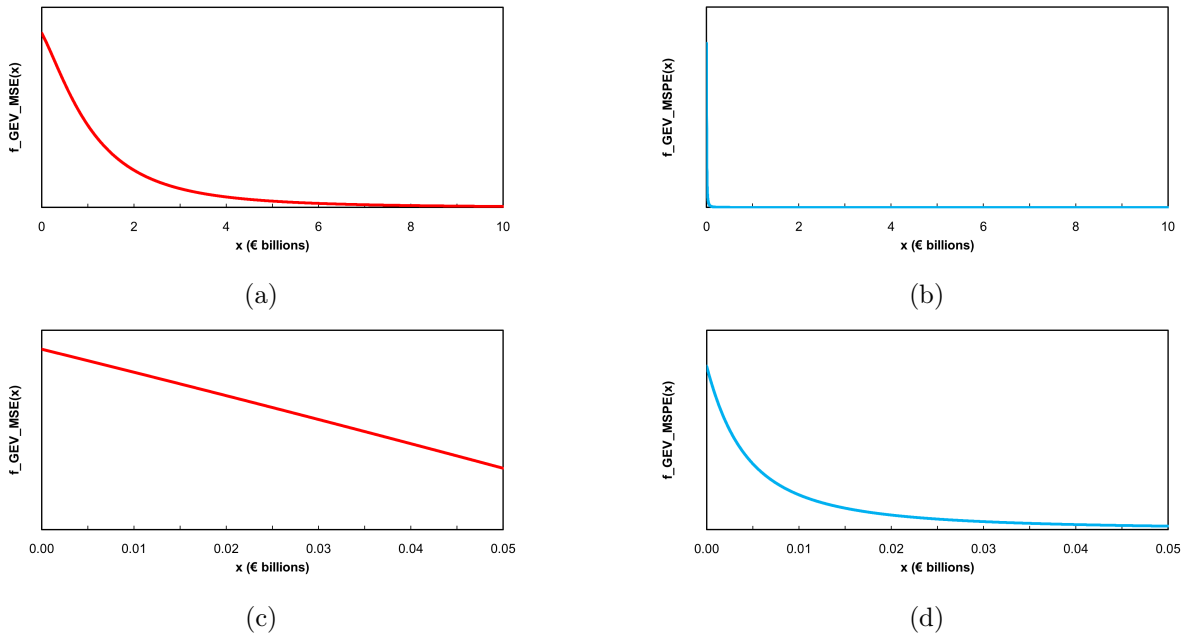


Figure 10: Movements of the PDFs of the MSE and MSPE distribution fits of the GEV distribution.

5.2.2 Truncation

Losses are always nonnegative. This means that the loss distribution fits should be truncated for $l \in \mathbb{R}_{\geq 0}$. Taking this into account in the quantile fitting procedure was only possible for the classic Student t-distribution by using the `crch` package in `R`. We obtain the other truncated distribution fits by plugging in the non-truncated fits and dividing the PDF of distribution i by $1 - F_i(0)$ so the PDF remains a proper density. For the non-truncated GEV distribution fits based on the MSE and MSPE error metrics, $F_{\text{GEV,MSE}}(0) = F_{\text{GEV,MSPE}}(0) = 0.368$. This implies that more than a third of the distribution lies below zero. However, we should further examine the segment of the distribution below zero.

First, we use results from Equation 7 regarding the domain of l to show that $F_{gev,\text{MSPE}}(l) = 0$ for $l \approx -0.00469$, which shows that negative values at worst in the order of 1 million in negative damages. Contrarily, $F_{gev,\text{MSE}}(l) = 0$ for $l \approx -2.032$, which means that this distribution occasionally produces negative billions in damages. Truncation is therefore crucial to obtain a realistic distribution for real-life losses, especially for the MSE fit. Table 5 shows the impact of truncating the distributions on the score and quantile value estimates. Notice the increasing scores for the truncated GEV distributions with the same variables as the non-truncated fits. An

explanation for this is that the distribution is mapped on the new interval $[0, \infty)$, which scales up every value that was already within this interval. The scores for the truncated GEV distributions proof to be a lot worse than the initial estimated truncated classic Student-t distribution. Hence, this naive truncation method after the parameter optimization is not accurate for this application. Truncation should be considered during the parameter optimization when available. Hereafter, we do not use the truncated GEV distributions within this thesis.

Table 5: Optimized loss distribution fits. The scores show the squared differences between the distribution fit quantiles and the actual proxy quantile values. The second row shows the actual damage proxy values for each quantile.

	Distribution	Score	Q0.9 (bn€)	Q0.99 (bn€)	Q0.999 (bn€)	Q0.9999 (bn€)
			0.0445	**11.1116**	**33.6455**	**87.1544**
MSE	ST	17.63	3.87	11.59	32.01	87.52
	GEV	**10.36**	3.09	11.42	32.67	87.37
	TGEV	417.58	4.20	14.22	39.87	104.91
MSPE	ST	1.78	0.0585	0.5899	5.8997	58.9974
	GEV	**1.48**	0.0486	0.6697	8.14	97.9345
	TGEV	2.70	0.0847	1.104	13.37	160.77

Notes: Scores can only be compared for the same error metric. For the Score column, the best score shows in bold for each error metric.

An alternative approach to make sure that losses are nonnegative at all times when simulating losses is by simulating typically from the non-truncated distribution fits for the GEV distribution and then setting negative values towards zero. This method seems feasible, since this adjustment has no influence on the right tail of the distribution. However, a consequence is that the value of zero increases substantially as more than one third of the values are set to zero. This affects the parameters as the location parameter shifts to the right and the variance decreases. However, a large peak at zero for the simulations does intuitively make sense, since losses due to flooding do not occur often and we consider the losses not conditional on default that is a flooding.

5.3 CAT Bonds

Now that we have explored the loss distribution fits, we consider the possibility of issuing CAT bonds for insurers as a reinsurance pool. For continuation and simplification purposes, again consider an insurer portfolio containing every residential building in South-Holland. We first examine the MSE fit for the non-truncated GEV distribution for the purpose of β^* having an analytical solution. Next, we simulate losses for the MSE fit for the non-truncated GEV distribution with negative values set to zero. We use simulation results to approach β^* numerically

by using the sample mean and variance.

5.3.1 Analytical Optimization

For the analytical optimization results, we consider solely the MSE fit for the non-truncated GEV distribution with CDF $F_{gev,\text{MSE}}(l)$. We use Equation (24) to calculate β^*, the optimal amount of CAT bonds to be issued according to the mean-variance hedge. Additionally, we investigate how CAT bond parameters and insurers' risk aversion affect β^*.

For the analytical optimization, set the following parameters for a CAT bond with maturity $T = 1$. Let the CAT bond price in billions of euros be $y = 0.001$ (corresponding to one bond being one million euros), the attachment point in billions of euros be $a = 33.65$, the damage proxy value of the 0.999 quantile. Moreover, let the premium paid to investors be $r = 0.05$, the risk aversion parameter be $\gamma = 1$, which indicates that the fixed, covariance and variance terms in Equation (24) receive equal weights. Lastly, let the payout fraction paid to investors in case of a catastrophic event be $p = 0.25$. Note The attachment point is set at the actual 0.999 quantile value, which means a 1 in 1000 year expected loss. β^* solves for $30,552$ bonds for the exact parameters considered above. We calculate and examine how β^* changes with respect to individual parameters changing, ceteris paribus. In Section 4, we already established the effects of both r and γ individually.

Additionally, we investigate combined effects of the bond parameters on β^*. We consider the above set parameters as the base case and provide ranges for which we show the course of β^* with respect to the parameters within the range. Table 6 provides a structured overview for the base values and ranges of bond parameters changing for the analytical optimization. When we perform calculations of β^* for a range of parameters, we use 100 values at equal intervals within the parameter range at hand.

Table 6: An overview for base values and ranges for different CAT bond parameters for a CAT bond with maturity $T = 1$.

	a (bn €)	p	r	γ
Base value	33.65	0.25	0.05	1
Range	$[11.11, 87.15]$	$[0, 0.5]$	$[0.01, 0.1]$	$[0.5, 2]$

Notes: Values of a are based on the estimated quantile values for the damage proxies.

We obtain results for every combination of two parameters changing. Remember that we assign the base value for other parameters that stay fixed. Since we have four parameters in

total, we have $\binom{4}{2} = 6$ combinations to consider. Results for all six combinations show in Figure 11. Observe that $\beta^* > 0$ for most parameter value combinations. Figure 11a shows the earlier described course for β^* when increasing γ, converging from zero to $\frac{-\text{Cov}(L,P_{\text{CAT}})}{y(1+r)\text{Var}(P_{\text{CAT}})}$, a positive number as expected.

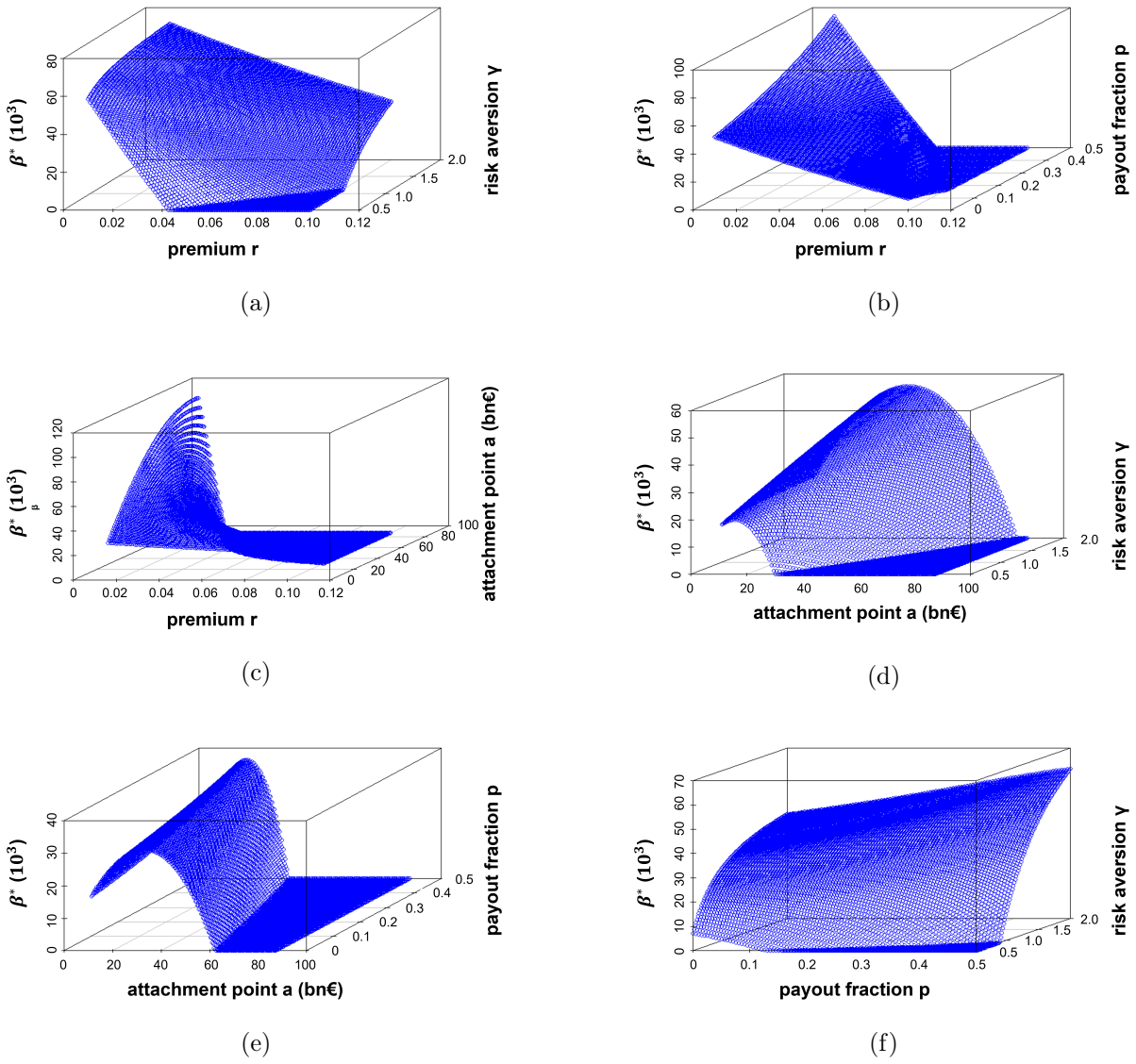


Figure 11: Graphs that show how various CAT bond parameters affect the optimal amount of CAT bonds to be issued, denoted as β^*. six combinations of combined parameter effects show. Parameters not considered in individual graphs have based values $a = 33.65$, $p = 0.25$, $r = 0.05$ and $\gamma = 1$.

Figure 11b confirms that effects of changing p on β^* depend on premium r, since r strongly influences the relationship between the expectation term and the covariance term in the nominator of β^*. For r close to zero, increasing p has a positive effect on the optimal amount of CAT bonds to issue. When r grows larger, the effect when increasing p slowly becomes negative. The effects of r and p on β^* depend on a variety of movements. First, increasing p increases the

expectation term, while decreasing the covariance term towards zero when p goes to one. Simultaneously, increasing p decreases the variance term, while increasing r decreases the variance term. It makes sense that for higher r and p, β^* goes to zero, since we choose $F_{gev,\text{MSE}}(a)$ close to one. If r increases, the expectation term becomes negative. For lower p, the covariance term may compensate, but when p increases, the covariance term shrinks towards zero, resulting in the denominator term in β^* to become negative. The movements also depend on the choice of γ and a.

Lastly, Figures 11c - 11e show a parabolic movement of β^* when a changes. The parameter range for a is set such that $0.99 \leq F_{gev,\text{MSE}}(a) \leq 0.9999$. When a increases such that $F_{gev,\text{MSE}}(a)$ approaches 1, β^* becomes zero, which confirms our derivations in Section 4.3.3. The parabolic movement of β^* when changing a within the parameter range is not immediately clear from Equation (24). An explanation for this movement is the relative differences in the terms in the nominator and denominator as both terms shrink with a different speed. The variance term includes the term $F_i(a)(1 - F_i(a))$, which could cause the parabolic movement, since the other effects of a on β^* are non-quadratic.

5.3.2 Optimization by Simulation

The analytical solution considers CAT bonds with maturity $T = 1$. Hence, we perform numerical optimization by simulation to expand this to higher maturities. We independently simulate $n = 10,000$ 1-year losses and corresponding P_{CAT} values and optimize Equation (15) with respect to β by using the sample mean and variance of P_{INS}. We perform the numerical optimization for the GEV distribution based on the MSE error metric, previously used in the analytical optimization. We obtain equal results, confirming our analytical derivation.

Additionally, numerical solutions allows us to set negative losses to zero and still perform optimization. As discussed, we use the non-truncated GEV distribution fit based on the MSE and set negative losses to zero. We use the simulated losses to retrieve the corresponding values of P_{CAT}. We perform the maximization for the mean-variance hedge numerically and retrieve $\hat{\beta}^*$ by using the sample variance and covariance in Equation (24).

We consider the base values from Table 6 and simulate aggregated losses for maturities 1 through 10, 15, 20 and 25. Results for β^* show in Figure 12. First, β^* moves up when increasing the maturity, but this number decreases at the later years. An explanation for this is the different magnitudes of increase for the variance and mean of P_{INS} due to β. Larger values of β combined with larger aggregated losses due the higher maturity result into a higher variance of P_{INS}, decreasing the optimal value of β^*.

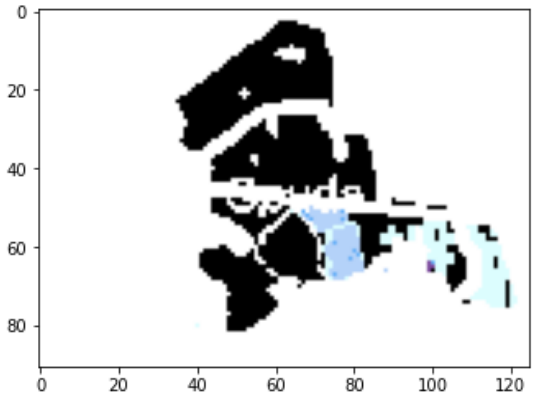
Figure 12: Optimal amount of CAT bonds to issue from the insurers' perspective for maturities 1 through 10.

5.4 Sensitivity Analysis: Zoom

For all flood scenario maps, we handle data by setting the zoom to 6. We now investigate whether results substantially differ for more detailed maps, corresponding to zoom 8. Zoom 8 corresponds to a 4 times more zoomed in scenario map. In line with previous visualization, we perform the sensitivity analysis for the municipality Gouda. An important note is that this analysis does not affect the underwriters' application from Section 5.1.2.

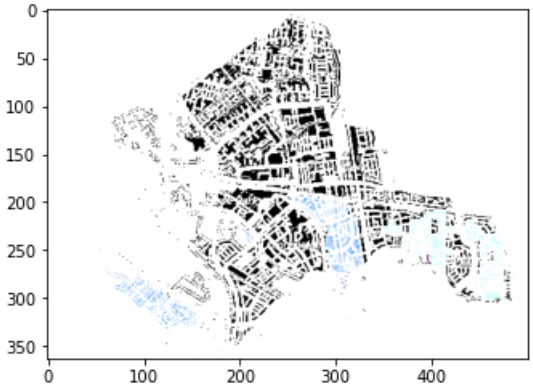
Figure 13 provides visualization of the differences when running the image recognition algorithm. The differences in detail of individual buildings is inherently clear. Additionally, the issue of the city name potentially affecting specific pixel positions becomes less important. Figure 13d shows that specific building areas do not show in more zoomed out flood scenario maps, while those areas are relevant for assessing flood risk.

(a) Zoom 6, building pixels.



(b) Zoom 6, 1 in 100-year flood scenario.

(c) Zoom 8, building pixels.



(d) Zoom 8, 1 in 100-year flood scenario.

Figure 13: Pixel data graphs of the municipality Gouda by using different zooms of flood scenario maps. Pixel data for both zooms include pixels that correspond to buildings and a 1 in 100-year flood scenario.

We convert the pixel data maps to damage proxies for all four flood scenario maps. Results for both zooms show in Table 7. While adjusting the zoom clearly alters the damage proxy estimates for Gouda, it matters less when we look at quantiles closer to one. Observe that increasing the zoom does not necessarily lead to under- or overestimation, since zoomed in results may either increase or decrease the damage proxies for different quantiles. Investing time in running the algorithm for scenario maps with higher zoom may be of interest when focusing on 1 in 10-year or 1-in-100 year floods.

Table 7: Damage proxy values for different quantiles for the municipality Gouda. Results include two different zooms. The results for zoom 8 are based on a four times more zoomed in flood scenario map opposed to the zoom 6 results.

Zoom	q0.9 (€)	q0.99 (€)	q0.999 (€)	q0.9999 (€)	Buildings	WOZ (€)
6	264,934	71,155,601	1,311,059,413	1,959,813,443	38,157	323,000
8	1,187,076	80,064,618	1,244,712,998	1,757,426,332	38,157	323,000

6 Conclusion

The focus of this thesis is to overcome the data scarcity within flood risk by extracting damage proxies and a loss distribution from existing flood scenario maps and to investigate the use of CAT bonds as a reinsurance pool for insurers to reallocate risk.

In particular, we propose a general framework to quantify flood scenario maps towards potential damages and obtain a loss distribution that forms the basis for the analysis of CAT bonds as a potential reinsurance pool. Recall the two research questions:

RQ1: *How can we extract and quantify damages to residential buildings from existing flood risk scenarios within the Netherlands?*

RQ2: *How can insurance companies use CAT bonds as a reinsurance pool for flood risk within the Netherlands?*

Regarding RQ1, this research proposes a novel image recognition algorithm to translate flood scenario maps to proxies for damages to residential buildings. We consider four flood scenarios corresponding to the 0.9, 0.99, 0.999 and 0.9999 quantiles for flood occurrences. We show the relevance of this procedure by assessing flood risk for insurer portfolios and a province South-Holland as a whole.

Moreover, we perform quantile fitting to retrieve a loss distribution. We find that the GEV distribution scores best when trying to replicate the actual quantile values, irrespective of the error metric. Concerning the error metrics, the MSE results into the best fit regarding the 0.99, 0.999 and 0.9999 quantiles, while neglecting the 0.9 quantile. The MSPE error metric results in the best fit for the 0.9 quantile, corresponding to more short term risk.

Concerning RQ2, this thesis finds that it is beneficial for insurers to issue CAT bonds as a reinsurance pool under most circumstances, even when a insurer is (not) risk averse. We therefore believe that CAT bonds are a viable option as a reinsurance pool within the Netherlands to reallocate flood risk. Furthermore, we provide an in-depth analysis on the optimal strategy for insurers regarding the issuing of CAT bonds. This analysis considers a broad range of bond parameter combinations and shows how changes in parameters affects the optimal amount of CAT bonds to issue. This thesis solely considers the insurers' perspective and we cannot make binding conclusions on CAT bonds being attractive from the investors' perspective.

Overall, this research provides a novel and flexible methodology to overcome data scarcity within flood risk and shows multiple applications for insurance companies or regulators to utilize knowledge regarding potential losses.

7 Discussion

This Section discusses assumptions, results and tackles some limitations of this thesis. Moreover, we propose possible further research.

7.1 Limitations

This thesis broadly shows the problems with data scarcity within the field of flood risk in the Netherlands. While we mainly propose a general methodology for obtaining damage proxies, this process shows evident limitations. First, the data preparation is time costly since screenshots require to be exactly aligned to the pixel for every flood scenario map. Flood scenario maps with high resolution downloads would solve this problem. Hereafter, we point out additional issues that result from the current flood scenario maps. City names in the overlay potentially result into incorrect assessment of buildings for specific pixels. The flood map shows no possibility to distinguish between residential buildings and other categories of buildings. Note that additionally, the flood scenario maps make the assumption that the scenarios do not alter between years. In reality, we could extend the analysis of damage proxies by researching new construction of buildings and future trends of WOZ or market values. This brings us to the next limitation that market values are not available for every residential building within the Netherlands, while this may prove a better proxy for the value of residential buildings.

The damage factors from Huizinga et al. (2017) show damage factors, specific to the Netherlands, for the insured sum at risk for different flood depths. However, the damage factor consider residential buildings, contents included, which makes the sole use of WOZ values as a damage proxy less correct. It is difficult to assess contents as content value may substantially differ per household. Note that aside from these limitations, this thesis still proposes a good base for a methodology for assessing potential losses. All of the above limitations involve changes in damage proxy estimates, but the order and course of damage proxies remains similar for alternative approaches.

An obvious limitation, also present due to data scarcity, is the use of four quantiles. The quantile fitting becomes very sensitive to the individual quantile values and our best distribution fit is therefore not necessarily universally the best choice in similar applications. Moreover, the quantiles all lie far within the right tail of the distribution. Quantile fitting analysis for the distributions does not consider the fact that for the real-world, yearly losses are zero very often. Draws for the distribution fit often output values close to, but not exactly zero. Additionally, four parameters limit the choice of distributions to distributions with a maximum of four parameters due to identification restrictions. None of the distribution fits adequately captured every quantile,

due to high differences in potential losses even far inside the tail. Truncation included in the parameter optimization was not possible within the scope of this thesis, since there is no currently available package to the best of our knowledge for the GPD, skewed Student t-distribution and GEV distribution. Implementation requires numerically solving the inverse CDF within every optimization step.

Furthermore, we limit ourselves to a fixed payout fraction p when exceeding the attachment point, while CAT bonds commonly use an additional exhaustion point as a parameter to indicate the maximum damage coverage of the contract, for which p is smallest (Stupfler and Yang, 2017). However, this only exerts a minor influence on the current CAT bond analysis in this thesis.

Although the amount of limitations is large, the general methodologies and analyses are fundamentally correct and provide a flexible framework for practical use. For example, the current formulas and analyses allow for alternative damage factors, market values, scenario maps, amount of quantiles and so forth.

7.2 Further Research

Many directions are possible for further research. We categorize further research into three categories: damage proxy estimation, distribution fitting and CAT bonds.

Regarding damage proxy estimation, this thesis provides a general framework and showcases the use of a image recognition algorithm for quantifying flood scenario maps. Practitioners may implement the algorithm for all of the Netherlands, with several zooms and with more sophisticated estimates for damage factors, market values. They may also include varying values for residential buildings over time via mapping of planned constructions and by including future value trends for residential buildings. Additionally, this thesis exclusively focuses on residential buildings, while many applications are possible related to other damage categories such as agricultural losses, infrastructure and industrial facilities.

For the distribution fitting procedure by using quantiles, it is possible to approach this problem by first using a probability of default for the occurrence of a flood with a distribution conditional on a flood event happening. Moreover, further research should include truncation for every distribution included in the parameter optimization.

In our approach for investigating the use of CAT bonds as a reinsurance pool, we exclusively look at the insurers' perspective. Additionally, it is interesting and relevant to also explore the investors' perspective to get better insights in realistic bond parameter ranges and whether issuing CAT bonds is still reasonable within those ranges. Research may extend this to the pricing of CAT bonds under risk neutral circumstances (Zonggang and Ma, 2013).

Lastly further research may look beyond mean-variance optimization problem that we extensively look at in this thesis. Extensions may distinguish positive or negative variance or even include the expected shortfall in the loss function. One can even expand the problem to a more dynamic multi-year setting, where insurers are able to issue CAT bonds each period. This dynamic setting would require machine learning methods to solve.

References

Adrian, T., Boyarchenko, N., and Giannone, D. (2019). Vulnerable growth. *American Economic Review*, 109(4):1263–89.

Armstrong, J. S. and Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International journal of forecasting*, 8(1):69–80.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(2):367–389.

De Myttenaere, A., Golden, B., Le Grand, B., and Rossi, F. (2016). Mean absolute percentage error for regression models. *Neurocomputing*, 192:38–48.

Dutch Association of Insurers (2020). Position paper overstroming: Hoe ga je om met het verzekeren van het restrisico tegen overstroming? https://www.verzekeraars.nl/media/8163/vvv-popa_overstroming_2020.pdf.

Dutch Association of Insurers (2022). Verzekeraars zetten volgende stap in verzekeren overstroming door grote rivier of zee.

Funda (2023). Funda. Accessed on: 9-9-2023 and 20-9-2023.

Hattermann, F., Huang, S., Burghoff, O., Willems, W., Österle, H., Büchner, M., and Kundzewicz, Z. (2014). Modelling flood damages under climate change conditions–a case study for germany. *Natural Hazards and Earth System Sciences*, 14(12):3151–3168.

Hodges, J. S. (2013). *Richly parameterized linear models: additive, time series, and spatial models using random effects*. CRC Press.

Huizinga, J., De Moel, H., Szewczyk, W., et al. (2017). Global flood depth-damage functions: Methodology and the database with guidelines. Technical report, Joint Research Centre (Seville site).

Jamalizadeh, A., Pourmousa, R., and Balakrishnan, N. (2009). Truncated and limited skew-normal and skew-t distributions: properties and an illustration. *Communications in Statistics-Theory and Methods*, 38(16-17):2653–2668.

Lee, J.-P. and Yu, M.-T. (2007). Valuation of catastrophe reinsurance with catastrophe bonds. *Insurance: Mathematics and Economics*, 41(2):264–278.

LIWO, Watermanagementcenter Nederland (2023). Liwo. https://basisinformatie-overstromingen.nl//maps.

Markowits, H. M. (1952). Portfolio selection. *Journal of finance*, 7(1):71–91.

Mitchell, J., Zhu, D., and Poon, A. (2022). Constructing density forecasts from quantile regressions: Multimodality in macro-financial dynamics.

Morrison, J. E. and Smith, J. A. (2002). Stochastic modeling of flood peaks using the generalized extreme value distribution. *Water resources research*, 38(12):41–1.

NOS (2022). Een jaar na de overstromingen in limburg wachten nog altijd mensen op hun huis. Accessed on: 9-9-2023.

Schweizer, M. (2010). Mean-variance hedging and mean-variance portfolio selection. *Encyclopedia of Quantitative Finance*, page 1177.

Statistics Netherlands (2023a). Dwellings; ownership, occupancy, region. Accessed on April 20, 2023 from CBS StatLine database (identifier 82900NED) https://opendata.cbs.nl/statline/portal.html?_la=en&_catalog=CBS&tableId=82900NED.

Statistics Netherlands (2023b). Gemiddelde woz-waarde van woningen op 1 januari; eigendom, regio. Accessed on April 20, 2023 from CBS StatLine database (identifier 82900NED) https://opendata.cbs.nl/statline/portal.html?_la=en&_catalog=CBS&tableId=82900NED. Gewijzigd op 16-2-2023 02:00.

Stupfler, G. and Yang, F. (2017). Analyzing and predicting cat bond premiums: a financial loss premium principle and extreme value modeling. *ASTIN Bulletin*, 48.

Wang, Q. (1991). The pot model described by the generalized pareto distribution with poisson arrival rate. *Journal of Hydrology*, 129(1-4):263–280.

Zonggang, M. and Ma, C.-Q. (2013). Pricing catastrophe risk bonds: A mixed approximation method. *Insurance: Mathematics and Economics*, 52:243–254.

Appendix A Alternative Damage Proxy Estimation

We use an alternative way for calculating damages when obtaining pixel data for the flood scenario maps. This involves another graph from Huizinga et al. (2017) where damages are measured by € per m^2, which shows in Figure 14. The flood scenario maps includes a legend for the scale. From this, we extract the real-world m^2 for one pixel in the data.

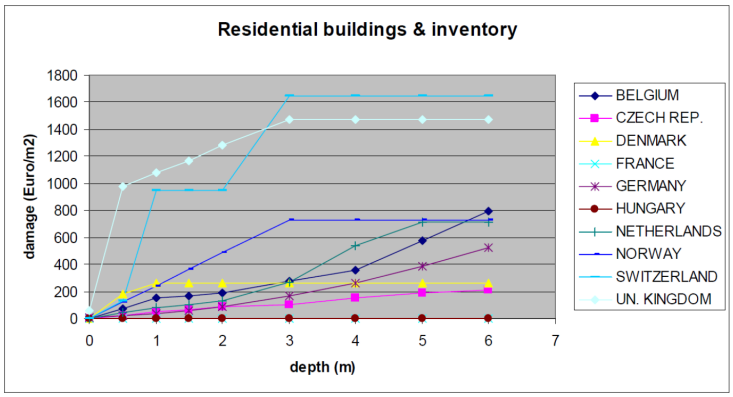


Figure 14: Depth-damage curves for a selection of European countries, gathered from Huizinga et al. (2017). Damages are considered per square meter for residential buildings and include inventory damages.

We use the (interpolated) damages per depth per m^2 from Figure 14 to estimate damage proxies from the pixel data after running the image recognition algorithm. Estimations for municipalities from Equation (28) change as follows:

$$\hat{L}_{(m,q)} = \sum_{d \in D} pix_{(m,d,q)} \cdot (dm^2)_d, \tag{28}$$

with dm^2_d the damage per m^2 for depth range d. Notice that x is replaced by pix since we use the number of pixels instead of the fraction. Results for the aggregated municipalities show in Table 8 for the province South-Holland. Both approaches result into estimates in the same order of magnitude. A remarkable result is that damages differ by approximately a factor two for every quantile. A possible explanation is that South-Holland has relatively high WOZ values compared to other provinces, which is not taken into account by the alternative damage estimation.

Table 8: Damage proxy estimates for different quantiles for South-Holland. The results include the use of damage factors and an alternative approach with damage per m^2.

Estimation Method	q0.9 (€)	q0.99 (€)	q0.999 (€)	q0.9999 (€)
Damage factors	44,508,911	11,111,566,335	33,645,496,434	87,154,426,199
Damage per m^2	26,061,117	5,938,708,344	17,584,573,394	42,619,701,585

Appendix B Damage Proxy Estimates

Table 9: Damage proxy estimates for every municipality in South-Holland. Aggregated results show for all of South-Holland.

Municipality	q0.9 (€)	q0.99 (€)	q0.999 (€)	q0.9999 (€)	Buildings	WOZ (€)
Vlaardingen	145121	17049652	487215829	1940501445	40817	277000
Goeree-Overflakkee	43312	10509658	36984213	1413113424	31388	347000
Voorne aan Zee	1367869	73974976	85690929	2277695825	38196	369000
Nissewaard	67376	307013786	875805326	3403686537	44451	279000
Hoeksche Waard	4271369	27753121	128752021	825648057	44517	361000
Dordrecht	12418434	17256376	22937231	5006848926	65629	297000
Gorinchem	488691	935047081	1777449963	1868336807	20017	312000
Hardinxveld-Giessendam	0	719949467	871785054	871785054	8515	335000
Sliedrecht	84525	533076636	945264527	935614344	12724	304000
Papendrecht	587613	1289615418	1596051032	1586103451	16571	332000
Hendrik-Ido-Ambacht	288289	21836717	21250557	449313226	14198	398000
Zwijndrecht	95463	7289733	20762742	640729601	23951	304000
Barendrecht	0	0	0	1307786812	22828	421000
Alblasserdam	895088	856291163	980760801	980609435	9756	325000
Ridderkerk	0	2909075	2909075	2284962922	25687	330000
Albrandswaard	0	0	0	648320227	11817	431000
Molenlanden	1488048	1829439328	2364265692	2433377505	20501	389000
Krimpen aan den IJssel	3000271	265137574	1518050601	1519094201	14366	354000
Capelle aan den IJssel	1117848	201958281	686729924	2298877788	34888	310000
Rotterdam	9229277	363029008	4487196138	17762443186	354192	320000
Westland	0	40806558	60618641	1987253312	55545	423000
Maassluis	0	13365279	15341523	542494902	18223	304000
Midden-Delfland	0	62934776	63082225	492360377	9113	469000
Delft	0	61111497	225332906	511547935	57386	322000
Schiedam	0	3088284	7222541	1286508982	41838	285000
Pijnacker-Nootdorp	0	42024952	46658695	61642657	24798	481000
Lansingerland	0	130187282	172338230	338430314	28058	479000
Krimpenerwaard	5212024	99459902	3191617013	3192955244	27766	388000
Gouda	264934	71155601	1311059413	1959813443	38157	323000
Waddinxveen	0	34583202	1277240823	1606223089	15652	370000
Zuidplas	1874432	401098897	1417028575	2191504525	22916	412000
Bodegraven-Reeuwijk	0	135053734	1031783069	1029872916	17982	432000
Zoetermeer	0	35686275	2494448629	2757322605	62552	357000
's-Gravenhage (gemeente)	1568927	40995519	150012867	8322925307	297930	355000
Rijswijk	0	40144049	42034283	128063610	32399	347000
Leidschendam-Voorburg	0	143673119	228089710	321788527	41737	380000
Voorschoten	0	36950373	38927004	77627891	13341	487000
Zoeterwoude	0	80387199	81015936	82011289	4803	454000
Alphen aan den Rijn	0	316709322	1819439901	1867205402	56520	360000
Nieuwkoop	0	421085603	1075088865	1066222133	14076	447000
Kaag en Braassem	0	598392448	1025589612	1008552808	14450	448000
Leiderdorp	0	12425856	69575660	107034641	14545	413000
Leiden	0	109214769	159921359	480301249	69853	390000
Oegstgeest	0	16175882	22114429	450278975	12204	563000
Wassenaar	0	0	3117853	311490958	13896	731000
Katwijk	0	0	5580784	1659165702	31274	419000
Teylingen	0	16393701	21539170	410953969	18853	459000
Lisse	0	590777628	580824038	795632109	12325	405000
Noordwijk	0	252151	1512906	1389166208	26263	489000
Hillegom	0	78295427	97478119	263226347	11443	403000
Total	44508911	11111566335	33645496434	87154426199	1960907	387800
Zuid-Holland_ave	60018387	14244213611	42330100659	1,02608E+11	1960907	387800