VERITAS NATIONAL DAM INSURANCE PROGRAM (VNDIP)

Team: Veritas Actuarial Solutions

TEE BOON LE
CHAN ZI XUAN
SUNWAY UNIVERSITY, MALAYSIA
FACULTY ADVISOR: MS. CHAN LAY GUAT

Table of Contents

Executive Summary 1
Section 1: Introduction 1
1.1 Definition and Explanation of Terms 1
1.2 Method and Assumptions on Data Imputations 8
1.3 Predictive Model Description 11
1.4 CIR Model for Inflation and Risk-Free Rate Prediction 11
1.5 The Need for a National Insurance Program 13
1.6 Program Metrics 19
2.1 Underwriting Criteria and Coverage Eligibility for Dam Insurance 19
2.2 Policy Features 25
2.2.1 Feature 1: Regular Inspection Coverage 25
2.2.2 Feature 2: Dam Upgrade Scheme 31
2.2.3 Feature 3: Economical Loss Coverage 36
2.3 Project Implementation Timeline 38
2.4 The Termination of Insurance Coverage 40
Section 3: Financial Result 42
3.1 The Determination of Premium 42
3.2 The Financial Projection 44
Section 4: Assumptions - Detailed Explanation 53
Section 5: Risks and Risk Mitigation Considerations 56
5.1 Risk Assessment 56
5.2 Sensitivity Analysis 59
Section 6: Data Limitations 63
Appendix: Data Imputation 66
Appendix: Gradient Boost Predictive Model (XGBoost) 70
Appendix: Hyperparameter Tuning 75
Appendix: Cox Ingersoll Rox Model 77
Appendix: Monte Carlo Simulation for Coverage Limit 83
Appendix: Underwriting 90
Appendix: The Computation of Annual Premium 97
Appendix: Monte Carlo Simulation for Dam Failure 103
References 104

Executive Summary

According to Carrieann (2024), the demand for dams has surged significantly over the past century. However, aging infrastructure and the increasing need for maintenance pose critical challenges. By 2050, a substantial number of people will rely on dams that have exceeded their intended operational lifespan (*Ageing Dams Pose Growing Threat: UN,* 2021). The 2023 Derna Dam collapse in Libya, which tragically claimed over 5,000 lives, serves as a stark reminder of the catastrophic consequences of dam failures and underscores the urgent necessity of proper maintenance (Norwegian Refuge Council, 2024).

As the saying goes, *"It is better to prevent than to cure."* Thus, the Veritas National Dam Insurance Program (VNDIP) has the goals of:

1. Comprehensive coverage for dam maintenance and upgrade, reducing long-term infrastructure risks.
2. Financial protection against economic losses due to dam failures, supporting recovery efforts.

This insurance framework is believed to enhance the infrastructural resilience of Tarrodan dams, mitigate financial risks, and safeguard Tarrodan's long-term sustainability goals, ensuring that its energy and water resources remain secure for future generations.

Section 1: Introduction

1.1 Definition and Explanation of Terms

Before we jump into this report, some of the terms and definitions must be defined clearly to avoid confusion.

1. Types of Dams and Introduction to Earthen Dams

Dams can be classified by the type of construction material used (Federal Emergency Management Agency, 2013). We have:

Concrete Dams	Arch, Buttress, Concrete, Gravity, Masonry, Multi-Arch, and Roller-Compacted Concrete.
Embankment Dams	Rock, Earth, or other earthen materials that are resistant to erosion.

This national insurance program will focus on the "Earth" type dams, which are the dominant type of dams in Tarrodan (*See Table 4: Distribution of Dams by Type in Tarrodan*).

Earthen dams are built by mechanically spreading and compacting suitable soils sourced from borrow pits or excavated on-site in successive layers. (In practice, "Earthen Dam" and "Earthfill Dam" are used interchangeably in the industry). Earthen dams are considered the most common type of dam built (The Constructor, 2018). An earthen dam is designed as a non-overflow section with a separate spillway, where the spillway is a structure that is used to control the release of water from a dam. Earthen dams are typically trapezoidal in shape and rely on their weight to hold back the force of water. The reasons behind the widespread use of earthen dams are:

- The foundation requirements are not as rigorous as other types of dams.
- The main construction materials are derived from the local soil, which helps to save the cost of construction.
- Easy to build, since no special machinery is required.

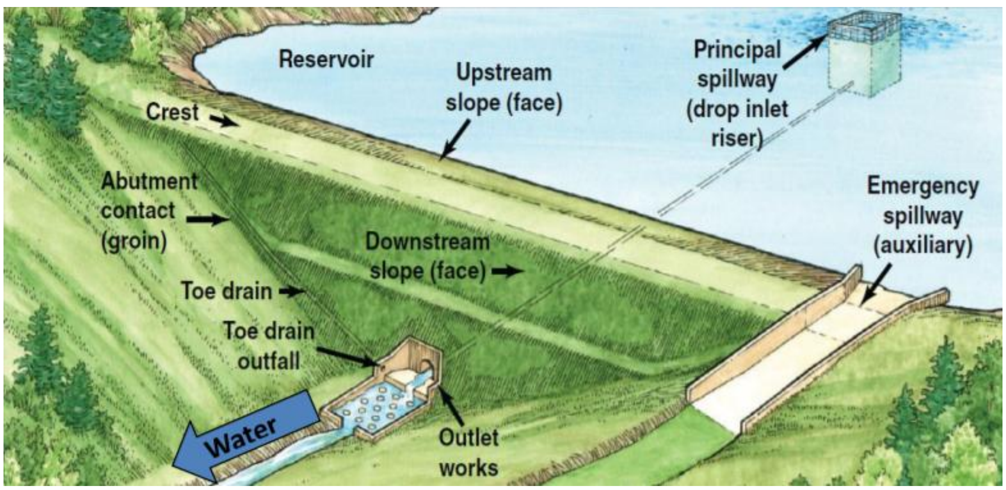


Figure 1: Red River Retention Authority. (n.d.). *Parts of a dam*. from https://www.redriverretentionauthority.net/parts-of-a-dam.html

To understand how an earthen dam works metaphorically, we can imagine an earthen dam as a sturdy bucket that is built to hold water, but it cannot let water spill over its sides. It has a separate drain designed specifically to handle overflow. Therefore, the main body of the earthen dam (the non-overflow section) is made from compacted layers of soil built to contain water, while any water exceeding the design level is safely directed away through a separate spillway. This dual system helps ensure that the dam itself will not be eroded or damaged by uncontrolled water flow.

Based on how earthen dams are constructed, we have:

Rolled Filled Earthen Dams	Successive layers of moistened or damp soil are laid one over the other. Each layer not exceeding 20 cm in thickness is properly compacted at optimum moisture level before adding the next layer.
Hydraulic Fill Earthen Dams	Hydraulic methods are used from the excavation of the soil to the construction of the dams. Soils are excavated and mixed with water to form a slurry, which is then pumped and poured along the dam's edges. As the slurry settles, the heavier, coarser particles drop out at the edges, while the lighter, finer particles flow toward the center. This natural process creates a watertight core in the middle without needing extra compaction.

Based on the material used to build an earthen dam, we have:

Homogeneous Dams	These dams are constructed with uniform and homogeneous materials, which make them simpler and less expensive to construct. These types of dams are used in areas where the expected loading stress is moderate.
Zoned-earth Dams or Non-Homogeneous dams	In contrast to homogeneous dams, a non-homogeneous dam is built with distinct layers or zones, each made from different materials, like a layered cake for higher stability. In regions with higher stress or more complex natural conditions, for example, seismic risks. Zoned-earth dams are chosen and built with specialized layers to manage those challenges.

Figure 2

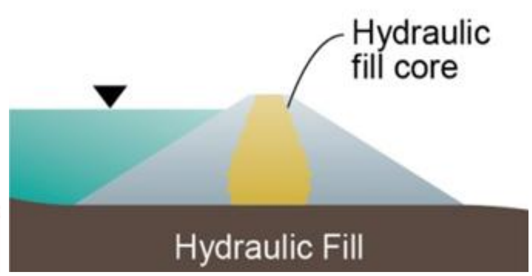


Figure 3

Figure 2: Cross-section of an earthfill dam with rock toe. Adapted from Federal Emergency Management Agency (2013).

Figure 3: Hydraulic fill dam section. Adapted from Federal Emergency Management Agency (2013).

2. Primary Purpose of dams

In Tarrodan, all Earthen dams have 11 different purposes (*See Table 5: Composition of Earthen Dams by Region and Purpose*).

Debris Control	Debris control dams are engineered to capture and manage floating debris such as logs, sediments, organic matter, and even trash that accumulates in streams or rivers over time (Hoellein et al., 2024). They reduce the risk of blockages and infrastructural damage or even flood risk (since sediments built up in riverbeds will reduce the capacity of the river channels, which in turn exacerbates flood risk). They are strategically constructed in areas prone to high debris flows, such as regions with dense forests experiencing heavy rainfall or rapid snowmelt.
Irrigation	Irrigation dams are built to ensure a reliable water supply for agriculture, improve water-use efficiency on irrigated land, and provide storage for spills, which is the excess water that overflows or escapes during irrigation, and tailwater, that is runoff that drains from fields after irrigation (United States Department of Agriculture, 2020).
Recreation	Recreation dams create outdoor recreational areas for activities like hiking, fishing, swimming, and boating. Meanwhile, it is vital for preserving aquatic habitats downstream and supporting biodiversity (ASDSO Dam Safety Toolbox, n.d.).
Flood Risk Reduction	Flood risk reduction dams are specifically designed to manage floods or minimize downstream damage by storing floodwater and gradually releasing it, thus lowering peak flows during periods of heavy rainfall. After initial testing (first filling and monitoring), many of these dams go years or even decades without a significant flood event, making it challenging to detect potential deficiencies in time to take preventive measures (ASDSO Dam Safety Toolbox, n.d.).
Fire Protection, Stock, Or Small Fish Pond	These dams play a critical role in regions prone to wildfires by providing a reliable, easily accessible water source for firefighting trucks or helicopters. Their presence also creates

	natural firebreaks, which can slow or even prevent the spread of wildfires. On regular days, these dams support aquacultural activities by serving as small fish ponds, contributing to both food production and recreational fishing opportunities (Hughson, 2023).
Water Supply	Water supply dams store and supply water for both domestic and industrial use. By ensuring a consistent water reserve, they help mitigate the impacts of drought or water shortages (ASDSO Dam Safety Toolbox, n.d.).
Hydroelectric	Hydroelectric is considered one of the oldest and largest sources of renewable energy. Thus, hydroelectric dams harness the gravitational potential of water stored at higher elevations; as water flows downward, it drives the turbines that generate electricity. This energy is then transmitted to a nearby power grid for distribution.
Tailings	Chemical substances are used to extract valuable minerals from rock ores, leaving behind a by-product known as tailings. Depending on the mining method, these tailings can be either liquid or solid. Typically, tailings are toxic or even radioactive. Apparently, tailings dams are built to store both the waste materials and the processed water. They need more maintenance and monitoring as they threaten the environmental health of the surroundings (IndustriALL, 2019).
Grade Stabilization	Grade stabilization dams are designed to secure steep or unstable slopes and embankments while reducing erosion. They incorporate engineered layers, geosynthetic materials, and effective drainage systems to prevent landslides or slope failures that could compromise the dam's structural integrity. These dams are particularly useful in areas with challenging terrain, ensuring that the dam remains stable and secure under various conditions (United States Department of Agriculture, 2020).

3. Hazard Classification

The definitions for hazard are borrowed from the Federal Emergency Management Agency (2013).

Table 1: Hazard Classification and Detailed Definition

Classification	**Definition**
High	Dams classified under this level are expected to result in the loss of human life and severely disrupt access to critical facilities in the event of failure or misoperation. The impact would extend to numerous public and private facilities, and the costs of mitigation could be extensive or even unfeasible.
Significant	Dams classified under this level are not likely to cause loss of human life but could lead to considerable environmental damage in the event of failure or misoperation. They may disrupt access to critical facilities and significantly affect major public and private infrastructures. Substantial mitigation efforts would likely be required for post-event.
Low	Dams classified under this level are not expected to cause loss of human life but low economic or environmental loss in the event of failure or misoperation. However, damage can be rapidly repaired. Losses are expected to be principally limited to the owner’s property.
Undetermined	The impact in the event of failure or misoperation is unknown.

4. Condition Assessment Classification

According to the Association of State Dam Safety Officials (2025), the condition assessment of a dam can be defined as follows:

Table 2: Dam Condition Assessment and Detailed Definitions

Classification	Definition
Satisfactory	No existing or potential safety deficiencies are recognized. Dams perform acceptably under all loading conditions (static, hydrologic, and seismic) along with the applicable regulatory criteria or tolerable risk guidelines.
Fair	No safety deficiencies are observed under normal loading conditions; however, rare or extreme hydrological and/or seismic events may expose a dam safety deficiency; further action is needed if the risk is deemed significant.
Poor	A dam safety deficiency is evident under loading conditions that are likely to occur; remedial actions are required. This rating may also be applied when the key parameters used to indicate potential dam safety deficiencies are uncertain, indicating that further investigations and studies are needed.
Unsatisfactory	A dam safety deficiency is recognized, immediate or emergency remedial action is required.
Not Rated	The dam has either not been inspected or it is not regulated. This rating is also applicable for the dams that have not been assigned a rating despite being inspected before.
Not Available	No data is available.

1.2 Method and Assumptions on Data Imputations

Due to the substantial amount of missing data in our original dataset, it is essential to apply suitable assumptions and advanced imputation techniques to generate a dataset that closely reflects real-life conditions. Before outlining our imputation assumptions, we first present a summary of the missing data in the provided dataset:

Column Name	Missing (%)
ID	0
Region	0
Regulated Dam	0
Hazard	0
Probability of Failure	0
Height (m)	0* (Note: Some of the entries are 0, which are not realistic and require replacement)
Loss given failure – prop (Qm)	0.03
Loss given failure – liab (Qm)	0.06
Primary Type	1.24
Primary Purpose	5.69
Year Completed	6.65
Assessment	12.19
Inspection Frequency	39.01
Assessment Date	46.97
Last Inspection Date	48.18
Distance to Nearest City (km)	49.16
Loss given failure – BI (Qm)	51.57
Spillway	61.45
Years Modified	91.29
Reinspection Indicator	Manually Added to the dataset
Modification Indicator	
Reassessment Indicator	

<u>Imputation Methodology and Underlying Assumptions</u>

Our team follows a conservative approach throughout the data imputation process, ensuring that each imputed value is as realistic as possible. The following summarizes our methodology and assumptions:

1. **Missing Assessments and Spillway Data:**
    - **Assessment:** All dams with missing assessment values are designated as "Not Available".
    - **Spillway Structure:** Dams lacking spillway structure information are assumed to be "Uncontrolled."
2. **Imputing Categorical Variables:**
    - **Primary Purpose:** Dams are grouped by Region, Hazard Classification, and Regulation Status. For each group, any dam missing primary purpose information is assumed to serve the most common purpose observed within that group.
    - **Primary Type:** Similarly, dams are grouped by Region, Hazard Classification, Regulation Status, and Primary Purpose. For each group, any dams missing primary type information are assumed to serve the most common type observed within that group.
3. **Imputing Numerical Variables:**
    - **Height (m):** Zero entries are unrealistic; such records are removed prior to imputation. The remaining missing heights are then imputed as the median height of dams within the same Region, Hazard Classification, Regulation Status, Primary Purpose, and Primary Type grouping.
4. **Multivariate Imputation via MICE:**
    - For complex categorical and numerical variables, we employ Multivariate Imputation by Chained Equations (MICE) in R. This method imputes missing continuous, binary, and categorical data using Fully Continuous Specification to ensure accuracy (RDocumentation, n.d.).
    - In particular, "Loss given failure – prop (Qm)", "Loss given failure – liab (Qm)", "Loss given failure – BI (Qm)" and other variables are imputed using the Predicted Mean Matching (PMM) method. Research indicates PMM preserves the distributional characteristics of the original data (*See University of Virginia Library, Getting Started with Multiple Imputation in R*).

5. Handling Date Variables and Creation of Indications:

- Variables such as "Last Inspection Date", "Assessment Date", and "Year Modified" contain mixed date and text formats. Direct application of MICE to these variables can lead to inaccuracies. Since these data are critical for determining whether the last inspection or assessment accurately reflects a dam's current condition, we have introduced indicator variables. These indicators are binary, taking the value "Yes" when an inspection, reassessment, or modification is deemed necessary based on the context and "No" otherwise.

- **Reinspection Indicator:** This indicator is set to "Yes" if any of the following conditions are met:

  I. No last inspection date or record is available.

  II. The sum of the last inspection date and the specified inspection frequency (i.e., the scheduled inspection interval) is earlier than January 1, 2024 (i.e., the record is overdue).

  III. In the absence of inspection frequency data, the last inspection date is earlier than January 1, 2019.

  ➢ Otherwise, the indicator is "No."

- **Reassessment Indicator:** Similar in concept to the Reinspection Indicator, this variable is set to "Yes" if:

  I. The assessment date is earlier than January 1, 2019.

  II. No assessment date is available.

  ➢ Otherwise, the indicator is "No."

- **Modification Indicator:** This indicator is set to "Yes" if:

  I. The dam is aged 50 or above and there is no recorded year of modification.

  II. The dam is aged 50 or above and a modification has occurred, but it was more than 50 years ago.

  ➢ The extent or "degree" of modification will depend on both the hazard classification and the dam's condition assessment (*See Policy Feature 2*).

1.3 Predictive Model Description

In this program, our team employed a gradient boosting machine learning framework called XGBoost, which is a method based on decision trees to model two critical outcomes for dam safety: the probability of dam failure and the total loss given failure. For both models, we incorporated a wide range of predictor variables, including dam physical characteristics (such as Height, Volume, Region, Age), risk-related attributes (such as Hazard classification, Assessment, Distance to the Nearest City) and operational features (such as Regulation status, Primary Purpose, Primary Type).

XGBoost operates by constructing an ensemble of decision trees, each trained on a subset of data. The trees are developed sequentially, with each subsequent tree focusing on correcting the errors of the previous ones. The final prediction is generated by combining the outputs of all trees in the ensemble through averaging (APMonitor Optimization Suite, n.d.). To avoid overfitting, we also implemented an early stopping feature (stop if no improvement after 30 rounds).

This model will be used to predict the probability of failure and the total loss after adjustment of key factors mentioned above to showcase the effectiveness of this insurance program [*See Appendix: Gradient Boost Predictive Model (XGBoost)*].

1.4 CIR Model for Inflation and Risk-Free Rate Prediction

In this program, our team used the Cox-Ingersoll-Ross (CIR) model, which is a widely used stochastic process for modeling interest rates and related financial quantities to predict two rates: the annual effective inflation rate and the 1-year risk-free annual spot rate from 2024 to 2033.

The reason for using the 1-year risk-free annual spot rate instead of the 10-year risk-free is because the shorter period rates are more sensitive to short-term changes in market conditions, which are better for short-term pricing and evaluation purposes (since this is a 10-year insurance program).

The CIR model is a calibrated stochastic process with mean reversion and non-negativity properties. It allows rates to fluctuate around a long-term average rate, theta (θ) at a speed determined by a constant, kappa (κ) while capturing market volatility. We will make use of the historical rates provided to calibrate our model to ensure accuracy.

Our teams use Maximum Likelihood Estimation to estimate the following parameters:

- **κ (kappa):** The speed of reversion toward the long-term mean rate.
- **θ (theta):** The long-term mean rate.
- **σ (sigma):** The volatility of the rate.

We use the theoretical distribution of the CIR model, a Non-Central Chi-Squared Distribution after proper scaling, to define a negative log-likelihood function. We then start fitting the distribution with some initial dummy parameter values and loop over each historical rate to compute the log-likelihood. A high penalty will be imposed if any parameters are non-positive, and finally, we optimize the likelihood function using the L-BFGS-B algorithm, which supports bound constraints (Kenton, 2023), to obtain the Maximum Likelihood estimates (MLE) for the parameters.

Once the parameters are obtained (*See Appendix: Cox Ingersoll Rox Model*), we use the function simulateCIR() in R to simulate the rates.

Table 3: Interest Rates

Year	Inflation Rate	1-Year Annual Risk-Free Spot Rate
2024	0.0228	0.0511
2025	0.0387	0.0706
2026	0.0221	0.0454
2027	0.0248	0.0446
2028	0.0323	0.0514
2029	0.0467	0.0683
2030	0.0483	0.0706
2031	0.0592	0.0849
2032	0.0354	0.0573
2033	0.0304	0.0489

1.5 The Need for a National Insurance Program

Table 4 shows that earthen dams constitute more than 96% of all dams in Tarrodan. Out of all 19,618 earthen dams, 11,367 dams are regulated. The rest of the 8,251 dams are not regulated. These dams play a crucial role in Tarrodan's water resource management, supporting purposes such as flood risk reduction, irrigation, and water supply (*See Table 5*). However, aging infrastructure poses a significant challenge, as many dams have been operated for decades (*See Table 6*).

In regions like Navaldia and Lyndrassia, most earthen dams (6,132 and 5,210, respectively) fall within the 50-to-100-year age range, with some exceeding 200 years. This indicates the need for proper maintenance and risk mitigation. Additionally, as seen in *Table 7*, a significant number of these aging dams have been classified under "High" or "Significant" hazard classifications.

For instance, Navaldia alone houses 1,501 high-hazard dams within the 50-to-100-year age range, with an average total loss of 871.13 Qm. Likewise, Lyndrassia records 1,107 high-hazard dams of the same age category, averaging a total loss of 538.51 Qm, while Flumevale follows closely with 641 high-hazard dams, amounting to an average loss of 824 Qm.

While *Table 8* shows that among these dams, the number of dams with "Reinspection Indicator" and "Modification Indicator" of "Yes" (meaning they have not been inspected for the past 5 years and need repair or rehabilitation) is 3,023. These dams are surely posing a large potential risk to the safety of the citizens and the economy of Tarrodan.

These figures highlight the urgent need for enhanced monitoring, regular maintenance, and strategic investment in dam safety programs to mitigate the risk of substantial economic losses.

Table 4: Distribution of Dams by Type in Tarrodan

Primary Type	Count
Arch	44
Buttress	73
Concrete	260
Earth	19,618
Gravity	394
Masonry	17
Multi-Arch	12
Other	111
Rockfill	221

Roller-Compacted Concrete	14
Stone	30
Timber Crib	12

Table 5: Composition of Earthen Dams by Region and Purpose

Region	Primary Purpose	Count
Flumevale	Debris Control	41
	Fire Protection, Stock, Or Small Fish Pond	111
	Fish and Wildlife Pond	121
	Flood Risk Reduction	393
	Hydroelectric	121
	Irrigation	978
	Other	178
	Recreation	224
	Tailings	14
	Water Supply	916
Lyndrassia	Debris Control	124
	Fire Protection, Stock, Or Small Fish Pond	2226
	Fish and Wildlife Pond	264
	Flood Risk Reduction	1198
	Grade Stabilization	651
	Hydroelectric	3
	Irrigation	1111
	Other	315
	Recreation	1983
	Tailings	26
	Water Supply	128
Navaldia	Debris Control	20
	Fire Protection, Stock, Or Small Fish Pond	569
	Fish and Wildlife Pond	36
	Flood Risk Reduction	2701
	Grade Stabilization	18

	Hydroelectric	16
	Irrigation	794
	Other	350
	Recreation	2795
	Tailings	57
	Water Supply	1136

Table 6: Composition of Earthen Dams by Region and Age

Region	Age	Count
Flumevale	0-50	533
	50-100	1694
	100-150	829
	150-200	41
Lyndrassia	0-50	2567
	50-100	5210
	100-150	243
	150-200	6
	200-250	3
Navaldia	0-50	1823
	50-100	6132
	100-150	496
	150-200	36
	200-250	4
	250-300	1

Table 7: Composition of Earthen Dams above Age 50 by Hazard Level

Region	Age	Hazard	Count	Average Total Loss (In Qm)
Flumevale	50-100	High	641	824.00
		Significant	212	525.57
	100-150	High	241	835.56
		Significant	129	684.01
	150-200	High	18	832.30
		Significant	4	397.58
Lyndrassia	50-100	High	1107	538.51
		Significant	228	476.93
	100-150	High	67	728.18
		Significant	26	501.86
	150-200	High	4	723.83
Navaldia	50-100	High	1501	871.13
		Significant	492	408.83
	100-150	High	242	942.51
		Significant	76	477.74
	150-200	High	19	956.08
		Significant	11	489.47
	200-250	High	1	528.10
		Significant	2	434.25
	250-300	Significant	1	254.50

Total loss is the sum of Loss given failure - prop (Qm), Loss given failure - liab (Qm) and Loss given failure - BI (Qm).

Table 8: Composition of Earthen Dams with Reinspection and Modification Indicator "Yes"

Region	Age	Hazard	Count	Average Total Loss (in Qm)	Median Probability of Failure
Flumevale	50-100	High	474	816.74	0.0958
	50-100	Significant	126	496.19	0.1285
	100-150	High	148	841.84	0.0986
	100-150	Significant	55	681.68	0.1321
	150-200	High	13	743.49	0.0976
	150-200	Significant	4	397.58	0.1312
Lyndrassia	50-100	High	819	492.52	0.1123
	50-100	Significant	208	479.42	0.1408
	100-150	High	46	690.91	0.1084
	100-150	Significant	24	492.68	0.1410
	150-200	High	2	309.55	0.1060
Navaldia	50-100	High	595	877.18	0.1050
	50-100	Significant	311	383.22	0.1373
	100-150	High	128	931.48	0.1122
	100-150	Significant	46	431.34	0.1430
	150-200	High	12	937.73	0.1156
	150-200	Significant	8	403.05	0.1264
	200-250	High	1	528.10	0.1130
	200-250	Significant	2	434.25	0.1331
	250-300	Significant	1	254.50	0.1334

A bubble plot would be helpful to visualize this dataset since there are many variables.

Figure 4: Distribution of High Hazard Dams Requiring Inspection & Modification Across Three Regions.

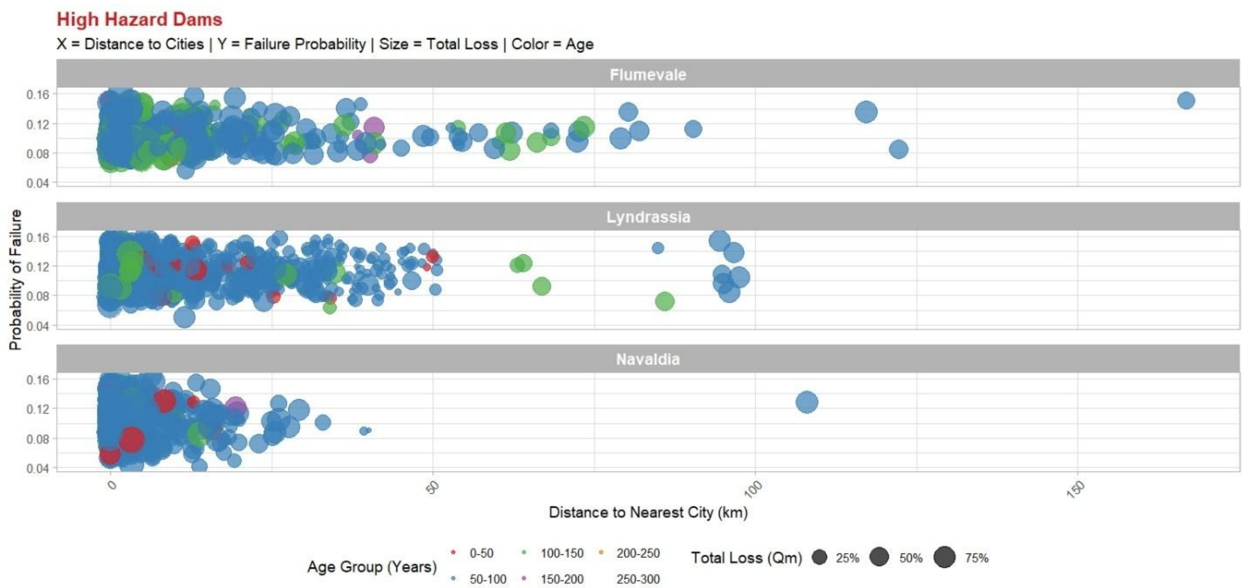
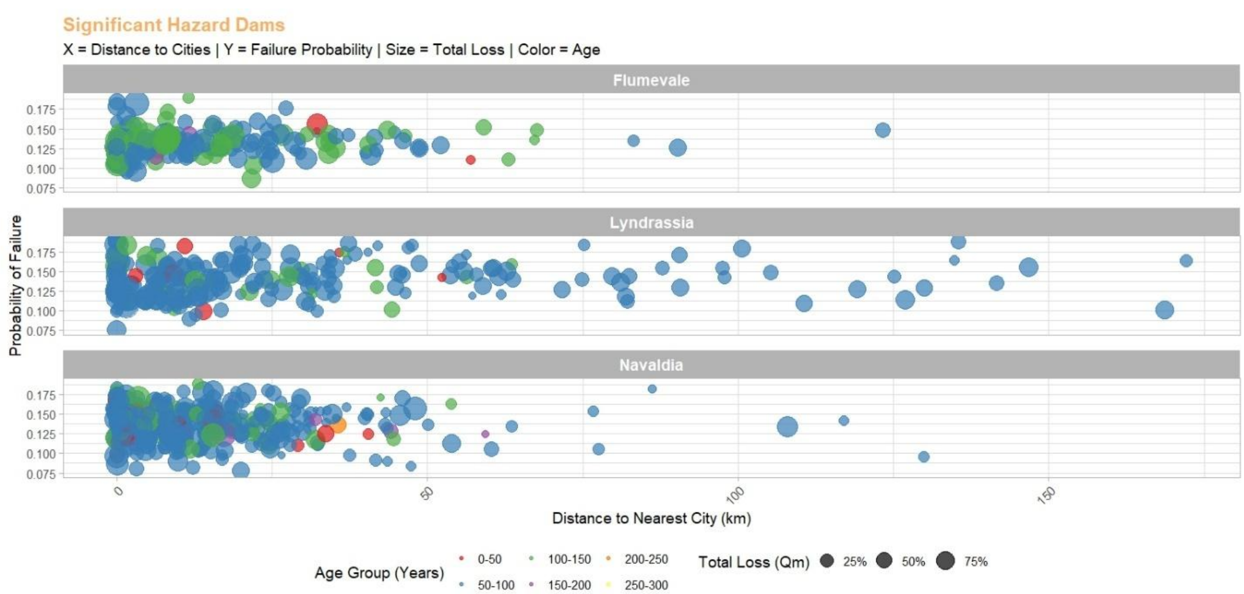


Figure 5: Distribution of Significant Hazard Dams Requiring Inspection & Modification Across Three Regions.



Note: The larger the bubble, the larger the total loss.

1.6 Program Metrics

The effectiveness of this national insurance program will be assessed based on the following key metrics:

1. **Government Commitment Level**: The extent of government participation, measured by budget allocations, policy endorsements, and legislative support for the insurance program.
2. **Claims Frequency and Severity**: Tracks the number of claims filed due to dam failures, along with the severity of losses, to assess the program's responsiveness to disasters.
3. **Payout Efficiency and Financial Relief**: Evaluates the speed and adequacy of claim payouts, ensuring timely financial support for affected regions to rebuild and recover.
4. **Risk Reduction and Hazard Mitigation**: Monitors the improvement of insured dams' safety after inspections and upgrades have been provided, ensuring a shift towards proactive risk management.
5. **Financial Sustainability**: Assesses the program's long-term viability by comparing premiums collected against claims paid out, ensuring financial stability.

Metrics will be reported on an annual basis, allowing necessary adjustments to be made so that the program remains effective.

Section 2: Program Design

The Veritas National Dam Insurance Program is a 10-year program that aims to provide comprehensive coverage to dam owners while effectively managing risk. Given the possible financial and environmental consequences of dam failures, the program includes strict underwriting standards and risk mitigation measures.

2.1 Underwriting Criteria and Coverage Eligibility for Dam Insurance

Our dam insurance program is designed to accommodate both regulated and non-regulated dams. To enroll, dam failure history records along with the most recent Hazard and Assessment report must be submitted to both our underwriting team and partnered engineering companies. These documents will be verified, and dams will then join the program after meeting the established criteria. (*See Appendix: Underwriting*).

Our underwriting process will incorporate an innovative rating system called the Hazard Index, where a higher index value indicates a greater risk faced by a dam.

$$Hazard\ Index = w_1x_1 + w_2x_2 + \cdots + w_{11}x_{11}$$

Table 9: Hazard Index Risk Zoning Categories

i	Description of x_i	Variable Value	Value of w_i
1	Regulated Dam	Regulated = 0; Not regulated = 1	1
2	Assessment Rating	Satisfactory = 0; Fair = 1; Poor = 2; Unsatisfactory/Not Rated/Not Available = 3	2
3	Hazard Classification	Low = 0; High = 1; Significant = 2; Undetermined = 3	2
4	Probability of Failure*	Low = 1; Medium = 2; High = 3	3
5	Age*	Low = 1; Medium = 2; High = 3	1
6	Total Loss*	Low = 1; Medium = 2; High = 3	3
7	Spillway	Controlled = 0; Uncontrolled = 1	2
8	Distance to the Nearest City (km)*	Far = 1; Near = 2; Close = 3	3
9	Modification Indicator	Yes = 1; No = 0	2
10	Reinspection Indicator	Yes = 1; No = 0	2
11	Reassessment Indicator	Yes = 1; No = 0	2

Note (*): The categorical values for numerical data (Probability of Failure, Age, Total Loss, etc.) are derived by splitting the data into three roughly equal groups, labeled as Low (1), Medium (2), and High (3).

Each dam will have its own index, and we will categorize them into 3 categories, namely: "High", "Medium", and "Low" according to the hazard index.

I. Regulated Earthen Dams

Criteria 1: Automatic Enrollment

Dams meeting **all** the following conditions are automatically enrolled:

- Hazard Classification: "Low," "High," or "Significant".
- Assessment Rating: "Satisfactory," "Fair," "Poor," or "Unsatisfactory".
- Reinspection Indicator: "No".
- Reassessment Indicator: "No".

*Among all 11,367 regulated earthen dams, 1,363 dams satisfy these requirements. We assume **100% of these dams join the program** for financial projection purposes.*

Criteria 2: Watchlist Enrollment

Dams that have the following conditions:

- Hazard Classification: “Low,” “High,” or “Significant”.
- Assessment Rating: “Satisfactory,” “Fair,” “Poor,” or “Unsatisfactory”.
- Either the Reinspection or Reassessment Indicator is “Yes.”.

are placed on a watchlist because the Hazard and Dam condition assessment may not be able to reflect the actual situation of the dams. They must undergo a regular inspection by our partnered engineering firms and provide the necessary documentation to verify their condition.

Among the 3,996 qualifying dams, we select those with a Hazard Index below the 90th percentile, resulting in 3,637 dams joining the insurance program.

Criteria 3: Conditional Enrollment After a Full Inspection

Dams in this group require a comprehensive, full inspection (*See Section: Types of Inspection*) by verified engineering companies to obtain updated ratings. There are two subcategories:

Option A:

- Hazard Classification: “Undetermined”.
- Assessment Rating: “Satisfactory,” “Fair,” “Poor,” or “Unsatisfactory”.

Option B:

- Hazard Classification: “Low,” “High,” or “Significant”.
- Assessment Rating: “Not Rated” or “Not Available”.

Among the 6,008 qualifying dams, we select those with a Hazard Index below the 70th percentile, resulting in 4,243 dams joining the insurance program.

Criteria 4: Reclassification as Non-Regulated

Dams with:

- Hazard Classification: “Undetermined”.
- Assessment Rating: “Not Rated” or “Not Available”.

are reclassified as non-regulated and will follow the corresponding underwriting criteria. No dams fall into this category.

II. Non-Regulated Earthen Dams

For non-regulated dams, applicants must provide a comprehensive **Technical Dossier** that details the dam’s construction, design, testing, and specifications. This dossier will be reviewed by the Tarrodan Dam Commission to confirm that the dam meets the basic safety requirements for an earthen dam. Based on the US Army Corps of Engineers (2004), a dam shall meet the following criteria (but not limited to):

- **Structural Integrity:**

  The dam must be designed to remain stable under all static and dynamic loads, with proper control of seepage and erosion. Efficient spillways and outlet systems are in good condition to prevent overtopping.

- **Maintenance & Monitoring:**

  A scheduled maintenance program (with proper documentation) must be in place, and the dam must be continuously monitored by trained operational personnel.

- **Emergency Preparedness:**

  Staff must be trained in emergency situations, and an effective alert system must be established.

- **Operational History:**

  A review of any past incidents, load capacity, and usage patterns is required.

- **Environmental Considerations:**

  The dam must comply with all relevant environmental regulations in Tarrodan.

Only dams that have passed the review of the Tarrodan Dam Commission can join the insurance program. We assume that 50% of the non-regulated dams can pass the review, which is 4,125 dams.

For dams that have passed the review, they need to meet the following criteria:

Criteria 1: Direct Enrollment

Dams that meet the following conditions can be directly insured:

- Hazard Classification: “Low,” “High,” or “Significant”.
- Assessment Rating: “Satisfactory,” “Fair,” “Poor,” or “Unsatisfactory”.
- Reinspection Indicator: “No”.
- Reassessment Indicator: “No”.

37 dams satisfy these criteria and join the insurance program.

Criteria 2: Watchlist Enrollment

Dams that have:

- Hazard Classification: “Low,” “High,” or “Significant”.
- Assessment Rating: “Satisfactory,” “Fair,” “Poor,” or “Unsatisfactory”.
- Either the Reinspection or Reassessment Indicator as “Yes”.

must undergo a regular inspection by dam safety personnel certified by our partnered engineering firms and provide the necessary documentation to verify their condition.

Among the 17 qualifying dams, we select those with a Hazard Index below the 90th percentile, resulting in 16 dams joining the insurance program.

Criteria 3: Conditional Enrollment After a Full Inspection

Dams must undergo a full inspection (*See Section: Types of Inspection*) with verified engineering companies to update their ratings before enrollment. There are two subcategories:

Option A:

- Hazard Classification: “Undetermined”.
- Assessment Rating: “Satisfactory,” “Fair,” “Poor,” or “Unsatisfactory”.

Option B:

- Hazard Classification: “Low,” “High,” or “Significant”.
- Assessment Rating: “Not Rated” or “Not Available”.

Among the 4,066 qualifying dams, we selected those with a Hazard Index below the 70th percentile, resulting in 2,853 dams joining the insurance program.

Criteria 4: Conditional Enrollment for Specific Cases

Dams with:

- Hazard Classification: “Undetermined”.
- Assessment Rating: “Not Rated” or “Not Available”.

can still join the insurance program after going through a full inspection by verified engineering companies, but we only select the dams that have a hazard index lower than the 50^{th} percentile among the group.

Among the 5 dams in this group, 3 are selected.

Under these selection criteria, the total number of insured dams is 12,152 dams.

<u>Prevention of Adverse Selection During Underwriting</u>

The prevention of adverse selection, where higher-risk dams may disproportionately join this insurance program is achieved through quantile-based selection. Dams are enrolled into the insurance program based on the quantile thresholds of the Hazard Index. This ensures that even among the eligible dams, enrollment prioritizes the dams with relatively lower risk so that the program maintains a balanced risk pool.

2.2 Policy Features

This comprehensive insurance policy aims to reduce long-term risk and provide coverage on economic loss from dams' failure.

2.2.1 Feature 1: Regular Inspection Coverage

Dams covered under this insurance program will benefit from scheduled inspections. Our approach incorporates a predictive model to capture the variability in the probability of failure associated with each variable in our dataset. We then apply the R function, partial(), to gain insight into how inspection will affect the predicted probability.

Figure 6 shows the relationship between the predicted probability of failure of a dam (y-axis) and the inspection indicator (x-axis). Each color represents a distinct hazard classification, with five samples collected for each category.

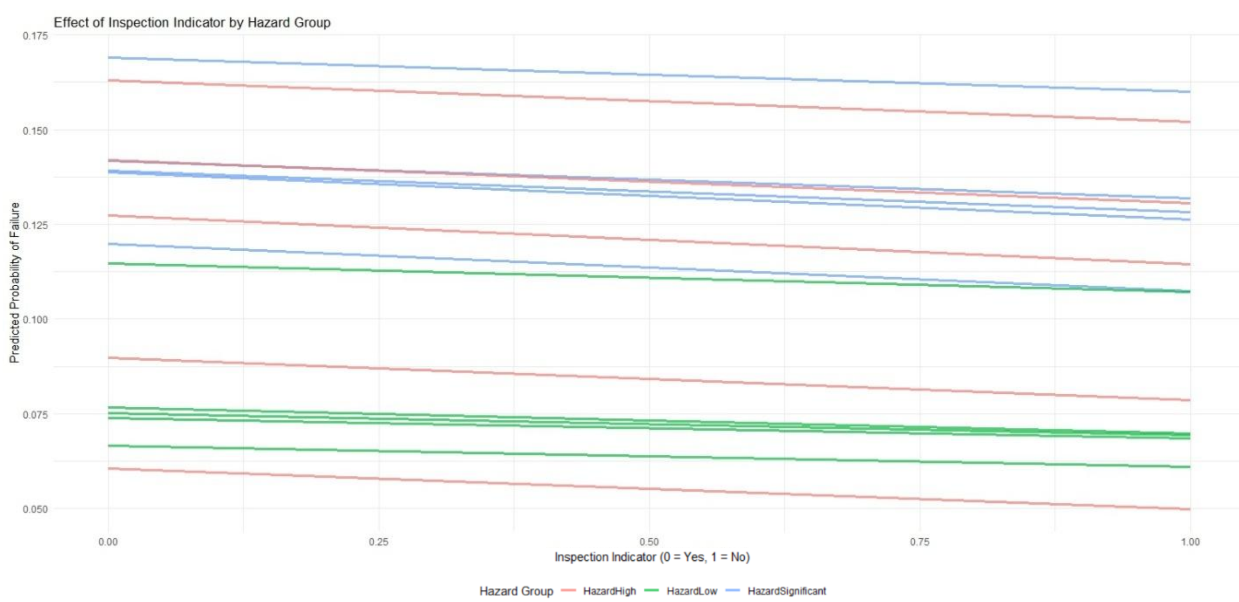


Figure 6: Effect of Inspection Indicator on Predicted Probability of Failure

It is evident that the predicted probability of failure is inversely related to the inspection indicator (reminder: "Yes" means inspection is required; the decreasing trend from "Yes" to "No" shows that the predicted probability of failure indeed decreased after inspection was performed), especially for hazard groups like "High" or "Significant" that indicate higher risk; the decreasing trend is more obvious. According to Department of Irrigation and Drainage Malaysia (2017), our team has developed comprehensive dam safety inspection standards. These standards will be proposed to our partnered engineering companies and implemented after review.

Types of Inspection and Inspection Guidelines

Note: While specific issues and their severity may vary by location, the following guidelines represent general practices for all earthen dams in Tarrodan.

The primary purpose of an inspection program is to detect dam safety deficiencies and symptoms of dam deterioration. Our team classifies inspections into different categories according to the personnel involved and purpose:

Types of Inspection	Personnel	Purpose
Full Dam Safety Inspection	Dam Engineers	Conduct a full-scale assessment to identify safety deficiencies by thoroughly reviewing all operational data and testing dam equipment to ensure full functionality and compliance with safety standards.
Routine Visual Inspection	Operational Personnel	Perform regular visual inspections of the dam's main structure to promptly identify and report any deficiencies.
Emergency Inspection	Dam Engineers	Reassess dam hazard classification following significant events (e.g., extreme floods, earthquakes) to determine the necessary remedial actions.

For insured dams, **the insurance program will cover the cost for a Full Dam Safety Inspection**. Before drafting an inspection guideline, it is important to identify the critical issues that need to be recognized and tracked during inspections.

1. Structural Elements: Examine the dam's body to identify the following issues:

What are the issues	Description	Consequences	How to detect	Recommended Actions
Cracks	Indicate potential foundation movement; it is often difficult to detect as they may be less than an inch wide yet several feet deep.	Early warning signs can eventually lead to slides: cracks can become weak points where water penetrates, accelerating internal erosion.	Visual Inspection: monitor changes in surface texture and document any widening or deepening of cracks.	Lower reservoir levels and have a qualified engineer assess the condition and recommend further remedial actions.
Slides	Represent significant slope failures, occurring when the forces on the dam's slope exceed the resistance of the building materials.	It can result in severe structural damage and obstruct the outlet.	Visual Inspection: look for early indicators such as bulges near the dam's toe or vertical displacements in the upper embankment.	

Sinkholes or Cave-ins	Arise from internal erosion of embankment materials, often following surface erosion by wave action or rain runoff.	It can reduce the embankment thickness, weaken its overall structure, and allow water to exit through the eroded area, potentially leading to dam failure.	Visual Inspection: look for signs of erosion on the embankment body, such as dirty water discharged at the dam exit.	
Seepage Areas	While all dams experience seepage, high-velocity seepage can indicate a potential problem, as rapid water flow through the dam body, it may lead to failure.	Excessive seepage can affect the dam's stability and eventually cause structural failure.	Visual Inspection: look for moss or marsh vegetation as they are indicators of persistent seepage when the reservoir is full.	

2. Spillway Structure: As noted earlier, the spillway acts as a safe outlet for excess water in the reservoir. Spillway inspection is an important part of the dam safety program.

What are the issues	Description	Consequences	How to detect	Recommended Actions
Obstructions	Caused by excessive vegetation growth (grass, weeds, brush, trees) or debris such as soil deposits from landslides.	An obstructed spillway will have a reduced discharge capacity, increasing the risk of dam overtopping. The dam body will have to endure excessive water pressure, increasing the chance of failure.	An obstructed spillway will have reduced discharge capacity, which can be detected by the sensors.	Periodic inspection and cleaning.
Erosion	Occurs when large volumes of water pass through the spillway during heavy rainfall. The spillway may be damaged or washed out completely.	May cause significant structural damage or obstruct the outlet.	Visual Inspection: look for unusual wear or displaced materials along the spillway channel.	Periodic inspection and maintenance.
Corrosion	Common in metal pipe spillways after being exposed to moisture or acidic conditions for extended periods.	May weaken the spillway structure and reduce water flow.	Visual Inspection: check for rust, pitting, or discolored water that could signal metal corrosion.	Use corrosion-resistant materials or keep metal pipes greased and painted.

3. Inlet / Outlet Structure: These structures may face similar issues like spillways, so regular inspection and testing are needed to ensure they are functioning properly.

Inspection Procedure

Note: The inspection procedures conducted by our partnered engineering companies must include, but are not limited to, the following items:

I. Pre-Inspection Preparation

- **Gather Documentation:** Collect all available records from previous inspections, maintenance logs, and engineering documents to establish a baseline for comparison.
- **Review Safety Protocols and Equipment:** Ensure every team member has appropriate Personal Protective Equipment (PPE) and standard operating procedures are in place for emergencies or unexpected findings.

II. On-Site Visual Inspection

- **Structural Elements**
  - Check for signs of cracks, slides, sinkholes, and seepage areas.
  - Check the water inlet and outlet.
  - Underwater structure checking.
- **Spillway Structure**
  - Assess whether side slopes have sloughed.
  - Check whether there is excessive vegetation in the channel.
  - Look for signs of erosion and rodent activity, sketch and photograph any parts with extensive erosion and record the width and depth of the erosion.
  - Examine moisture levels and hardness of the dam surface; document any areas that are unusually wet or soft.
  - Verify that the stilling basin (used to dissipate energy during spillway discharge) is properly protected with rocks or other armoring.

- **Inlet / Outlet Structure**
    - Functional testing of the inlet and outlet system components like valves or gates.
    - Check the backup power supply system.

III. Instrumentation & Data Collection

- **Check Monitoring Equipment:** Ensure piezometers, inclinometers, and other sensors that record dam performance are functioning properly.
- **Equipment Testing:** Ensure all equipment is in good condition and meets the safety requirements required by the Earthen Dam Commissions, especially parts that are submerged underwater.
- **Record and Comparison:** Compare current measurements, for example, water pressure and slope angles, with historical data to detect potential hazards.

IV. Documentation & Reporting

- **Inspection Forms:** Use standardized templates to maintain consistent records for future comparisons.
- **Organized Records:** Label and store photos and notes systematically so they remain accessible.
- **Remedial Actions:** If safety concerns arise, inform relevant authorities so that rehabilitative work can be planned and executed promptly.
- **Follow-up Inspections:** Schedule additional checks to monitor any issues identified and confirm that corrective measures are effective.

Dams are designed to retain water, so they will need to withstand massive hydrostatic pressure; any minor deformations should be given attention. After inspection, necessary actions shall be taken (*See Section 2.4: The Termination of Insurance Coverage*).

Inspection Frequency and Cost

Based on hazard classification, the inspection frequency for Full Dam Safety Inspection in this insurance program is determined according to the Department of Irrigation and Drainage Malaysia (2017):

Classification	Inspection Frequency	No. of Insured Dams
High	Annually	5246
Significant	Every 2 years	3763
Low	Every 5 years	3143

Dam inspection costs are not usually publicly available, as they are classified as confidential business information. The rates here are taken from the Montana Department of Natural Resources & Conservation Dam Safety Program (2025):

- The engineer's billable rate averaged $180/hour
- Visual inspection – typically 2 visits totaling 30 to 70 hours, costing around $12,600
- Evaluation and report writing – typically require 50 to 80 hours, costing around $14,400
- The approximate total average cost of inspection is around $27,000 per dam

Using the latest exchange rate (1 Qalkoon (Q) = 1.048 U.S. Dollar), the average cost of inspection is 25,763.35 Q per dam.

Considering inspection cost may vary according to a dam's complexity, our team will introduce loading factors to adjust the approximated cost of inspection. According to the hazard classification, the adjusted inspection cost is shown below.

Table 10: Approximated Cost of Inspection per Dam

Hazard Classification	Loading Factor	Approximated Cost (Base Rate times Loading Factor)
Low	1	25,763.35 Q
Medium	1.15	29,627.85 Q
High	1.3	33,492.35 Q

2.2.2 Feature 2: Dam Upgrade Scheme

As mentioned above, many dams have been working for decades; this undoubtedly poses significant risks. This insurance program is set to provide different types of upgrades to insured dams according to their ages and conditions if needed.

Dam upgrades are needed so that dams can meet the current safety standards. The inspiration behind this idea is that in the United States, approximately 65% of dams are privately owned, and the high cost of dam rehabilitation seems unaffordable for them (Association of State Dam Safety Officials, 2025). Therefore, this insurance program plans to provide repair, retrofit, and rehabilitation according to the dam's age and condition assessment classification.

The explanations incorporate insights from the Department of Irrigation and Drainage Malaysia (2017) and Central Water Commission (2018).

Criteria	Dam Repair	Dam Retrofit	Dam Rehabilitation
Objective & Definition	Targeted remedial actions to address localized deficiencies and ensure safe functionality.	Integration of modern technologies into existing dam structures to comply with current safety and operational standards.	A comprehensive strategy combining repair and retrofit measures to restore and enhance the overall performance, safety, and longevity of the dam.
Scope and Approach	Focus on specific defects or aging components. Common actions include replacing deteriorated electrical components (e.g., control panels, SCADA* systems, and sensors) that have reached the end of their intended service life. This includes patching concrete, sealing cracks, and addressing minor structural issues.	Involves upgrading and modernizing selected systems. Incorporates the installation of advanced instrumentation, updated control systems, and seismic improvements. Enhances existing drainage and monitoring systems without major structural changes.	Encompasses a full-scale overhaul of the dam's infrastructure. Combines targeted repairs and retrofits with extensive structural strengthening (e.g., foundation stabilization, spillway capacity improvement). Integrates the replacement of obsolete mechanical and electrical systems with modern, high-performance alternatives.

Note (*): SCADA stands for Supervisory Control and Data Acquisition.

According to the age and condition classification, we can divide all the insured dams into many categories; each category will receive different services.

Age	Condition Classification	Service
Age below 50	Poor	Repair
	Unsatisfactory	
Equal and Above 50 (With Modification Indicator of "Yes")	Fair	Retrofit
	Poor	
	Unsatisfactory	Rehabilitation
Equal and Above 50 (With Modification Indicator of "No")	Poor	Repair
	Unsatisfactory	

As for the estimated cost, we have retrieved the following figures from the Association of State Dam Safety Officials (2025).

Table 11: Cost of Dam Upgrades

Category (According to the height of the dam)	Repair	Retrofit	Rehabilitation
1 (Less than 4.5m)	$400,000 (381,679 Q)	$1,380,000 (1,316,794 Q)	$2,870,000 (2,738,550 Q)
2 (Between 4.5m and 7.6m)	$790,000 (753,817 Q)	$1,890,000 (1,803,435 Q)	$2,670,000 (2,547,710 Q)
3 (7.6m and 15.2m)	$1,410,000 (1,345,420 Q)	$4,000,000 (3,816,794 Q)	$6,230,000 (5,944,656 Q)
4 (15.2m and 30.5m)	$1,360,000 (1,297,710 Q)	$4,800,000 (4,580,153 Q)	$8,580,000 (8,187,023 Q)
5 (30.5m and 61m)	$3,080,000 (2,938,931 Q)	$20,000,000 (19,083,969 Q)	$23,840,000 (22,748,092 Q)
6 (Above 61m)	$9,180,000 (8,759,542 Q)	$26,340,000 (25,133,588 Q)	$95,300,000 (90,935,115 Q)

Under this feature, 6,375 insured dams are assigned a service based on their age group, assessment classification, and height category, as summarized in *Table 12* below.

Table 12: Distribution of Insured Dams by Age, Assessment, and Service

Age	Assessment	Service	Height	Count
50 and above	Poor	Repair	1 (Less than 4.5m)	10
50 and above	Poor	Repair	2 (Between 4.5m and 7.6m)	11
50 and above	Poor	Repair	3 (7.6m and 15.2m)	17
50 and above	Poor	Repair	4 (15.2m and 30.5m)	3
50 and above	Poor	Repair	5 (30.5m and 61m)	3
50 and above	Unsatisfactory	Repair	1 (Less than 4.5m)	6
50 and above	Unsatisfactory	Repair	2 (Between 4.5m and 7.6m)	10
50 and above	Unsatisfactory	Repair	3 (7.6m and 15.2m)	30
50 and above	Unsatisfactory	Repair	4 (15.2m and 30.5m)	5
50 and above	Fair	Retrofit	1 (Less than 4.5m)	284
50 and above	Fair	Retrofit	2 (Between 4.5m and 7.6m)	730
50 and above	Fair	Retrofit	3 (7.6m and 15.2m)	1118
50 and above	Fair	Retrofit	4 (15.2m and 30.5m)	261
50 and above	Fair	Retrofit	5 (30.5m and 61m)	35
50 and above	Fair	Retrofit	6 (Above 61m)	14
50 and above	Poor	Retrofit	1 (Less than 4.5m)	201
50 and above	Poor	Retrofit	2 (Between 4.5m and 7.6m)	510
50 and above	Poor	Retrofit	3 (7.6m and 15.2m)	694
50 and above	Poor	Retrofit	4 (15.2m and 30.5m)	81
50 and above	Poor	Retrofit	5 (30.5m and 61m)	11
50 and above	Poor	Retrofit	6 (Above 61m)	2
50 and above	Unsatisfactory	Rehabilitation	1 (Less than 4.5m)	188
50 and above	Unsatisfactory	Rehabilitation	2 (Between 4.5m and 7.6m)	488
50 and above	Unsatisfactory	Rehabilitation	3 (7.6m and 15.2m)	648
50 and above	Unsatisfactory	Rehabilitation	4 (15.2m and 30.5m)	85
50 and above	Unsatisfactory	Rehabilitation	5 (30.5m and 61m)	2
Below 50	Poor	Repair	1 (Less than 4.5m)	2
Below 50	Poor	Repair	2 (Between 4.5m and 7.6m)	4
Below 50	Poor	Repair	3 (7.6m and 15.2m)	7
Below 50	Poor	Repair	4 (15.2m and 30.5m)	1
Below 50	Poor	Repair	5 (30.5m and 61m)	1
Below 50	Unsatisfactory	Repair	1 (Less than 4.5m)	1
Below 50	Unsatisfactory	Repair	2 (Between 4.5m and 7.6m)	3

			3 (7.6m and 15.2m)	1
			4 (15.2m and 30.5m)	84
			6 (Above 61m)	352

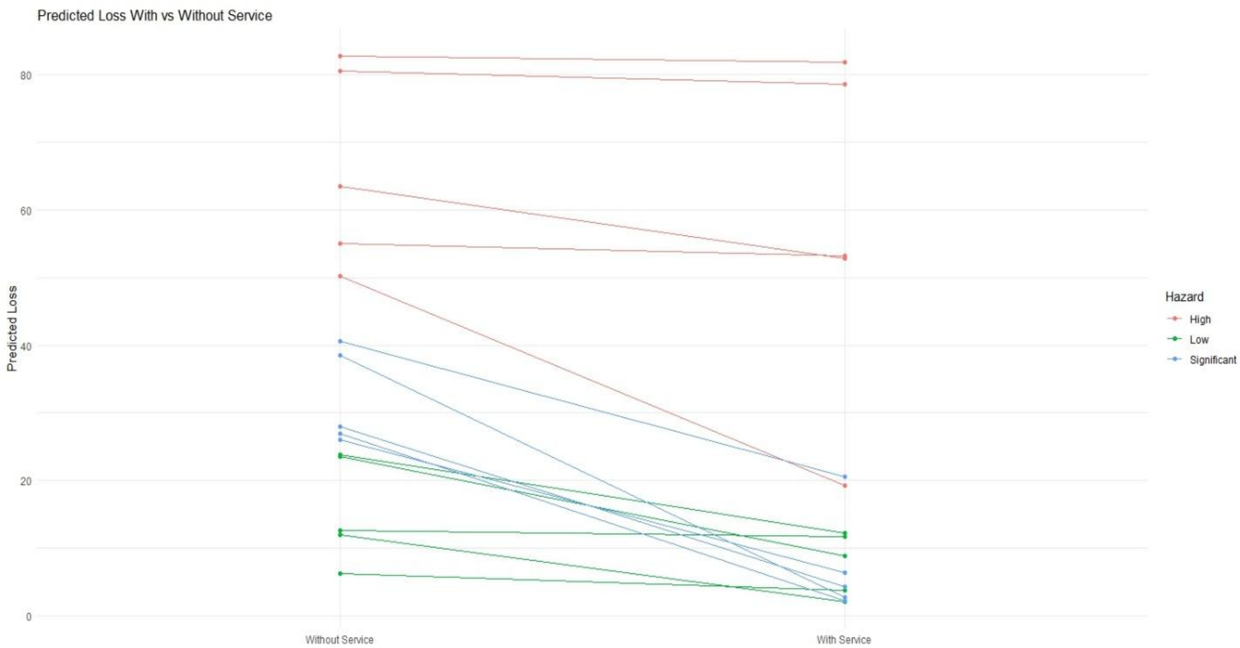


Figure 7: The Comparison of Predicted Total Given Failure with and without the Upgrade Service Provided

Figure 7 illustrates the changes in predicted total loss given failure for the same dams under different hazard classifications after the necessary service upgrades have been implemented. Five dams from each hazard classification were randomly selected for visualization.

The results clearly show that almost all dams have a decreasing trend in the predicted total loss given failure. Especially for dams in the "Low" and "Significant" hazard classifications.

The predicted total loss given failure was obtained using the predictive model XGboost. We assume that the upgrading services provided will improve the classification of dam hazards as well as their condition assessment.

A scatter plot will be used here to showcase how the insurance features help to achieve the objective of reducing long-term infrastructural risk. To avoid the plot being overcrowded, we will randomly pick 4000 individuals from the 12,152 insured dams.

Figure 8: The Comparison of Probability of Failure, With and Without the Insurance Program

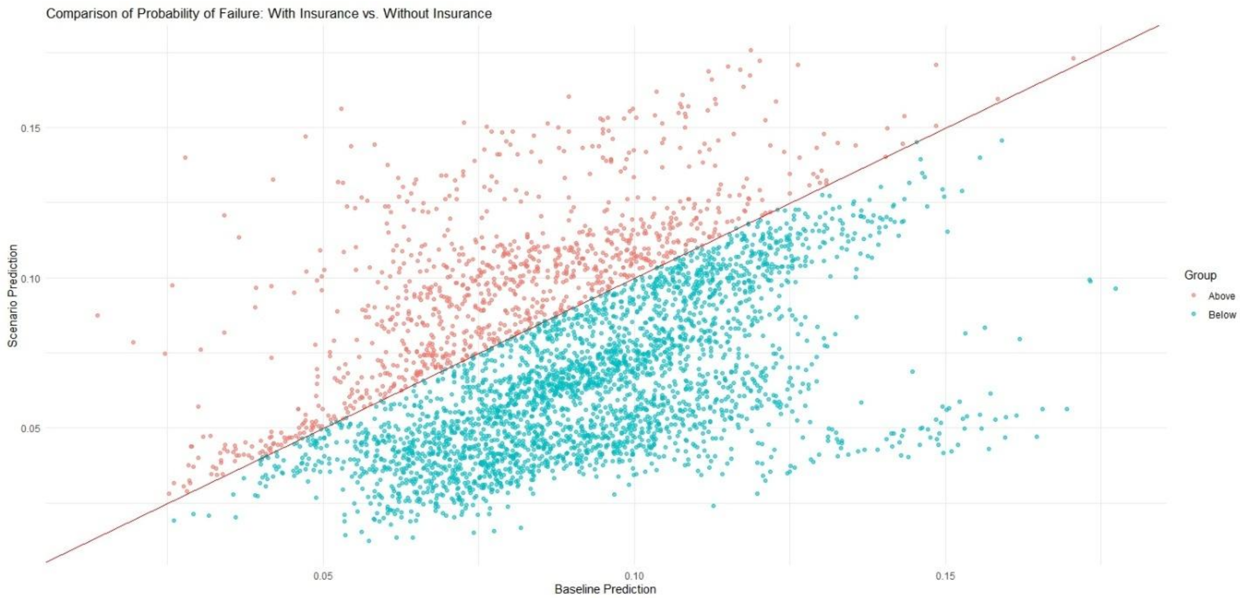
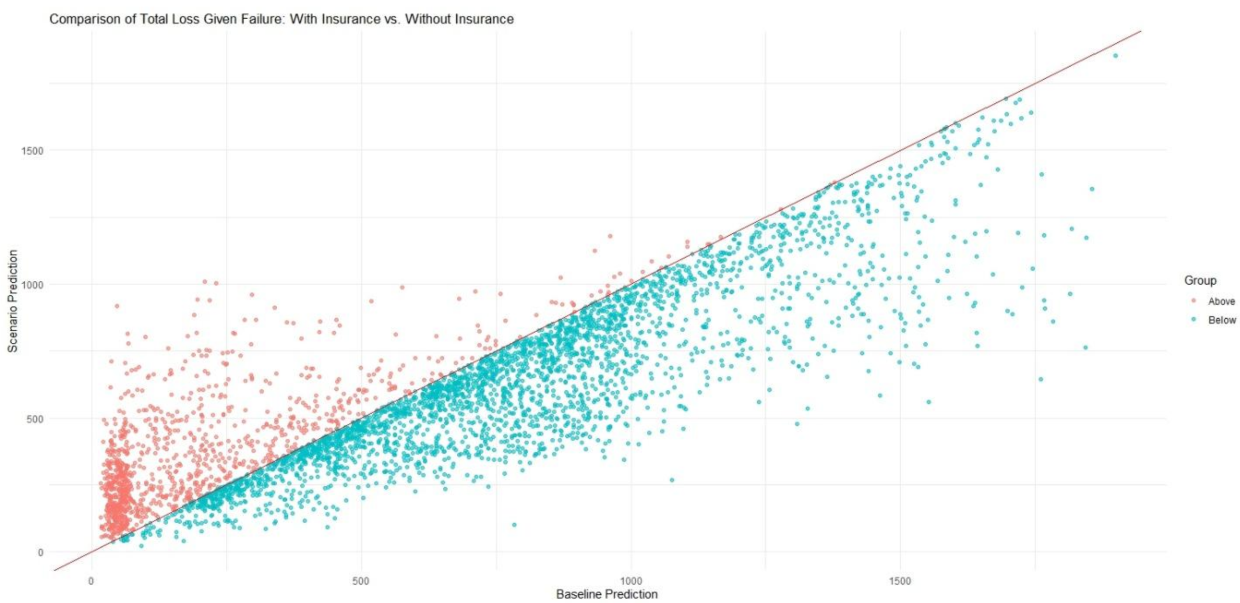


Figure 9: The Comparison of Total Loss Given Failure, With and Without the Insurance Program



It is evident that the insurance features help to achieve the goal since most dots are below the baseline, indicating a reduction in either the probability of failure or total loss given failure. The effects of outliers above the baseline will be studied.

2.2.3 Feature 3: Economical Loss Coverage

Structure of Coverage

According to the Hazard Index, each dam is classified into 3 categories, namely "Low", "Medium", and "High". Each group will have a similar number of dams; this is achieved using the cut_number() in R. The percentage of coverage for each category is different.

For category "Low", 100% of the loss is covered; for category "Medium" 80% of the loss is covered; for category "High" 60% of the loss is covered.

Table 13: Percentage of Coverage for Each Hazard Index Category

Category	Percentage of Loss Coverage
Low	100%
Medium	80%
High	60%

We are going to introduce Liability Sharing Structures between our insurance company and our reinsurance partners.

The liability sharing threshold is determined using Monte Carlo simulation along with modern machine learning, the XGBoost model previously constructed. We use Monte Carlo simulation to randomly assign dam characteristics such as "Height (m)", "Volume (m3)", "Primary Purpose" and so on, sampled directly from the original dataset to dams in the simulated dataset. After that, the XGBoost model is applied on the simulated dataset to estimate the total loss given failure for each dam (*See Appendix: Monte Carlo Simulation for Coverage Limit*).

For groups of dams sharing the same region and primary purpose, we take the 70th percentile of the simulated total loss given failure as the threshold. This means that 70% of the simulated outcomes fall below this threshold, providing a sound balance between risk retention and protection.

Below is *Table 14,* which summarizes the coverage limits (in Qm) for different types of earthen dams, categorized by region and their primary purpose.

Table 14: Coverage Limit for Earthen Dams from Different Regions and Purposes

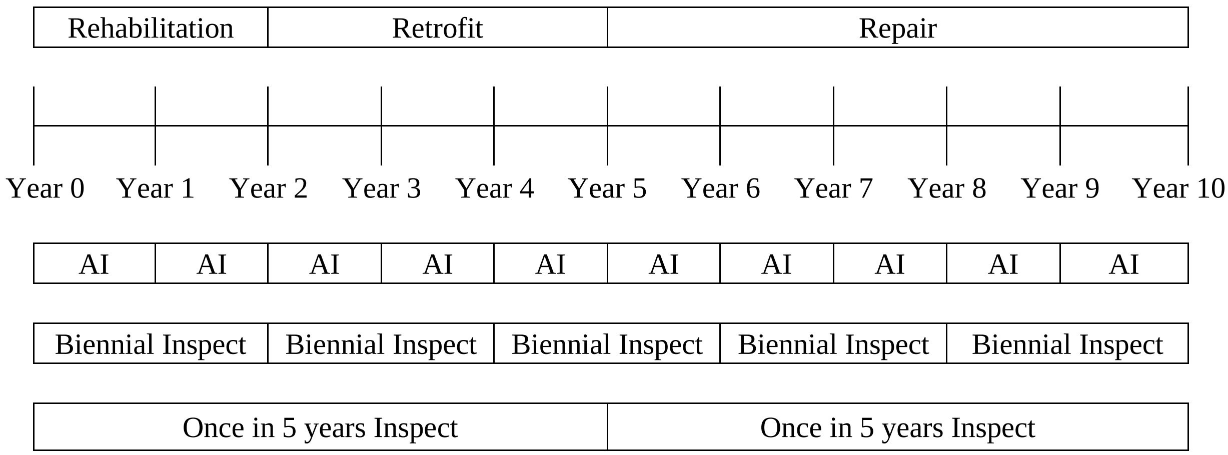
Region	Primary Purpose	Coverage Limit (Qm)
Flumevale	Debris Control	639.93
	Irrigation	606.16
	Recreation	596.57
	Flood Risk Reduction	586.49
	Fire Protection, Stock, Or Small Fish Pond	598.50
	Fish and Wildlife Pond	581.54
	Water Supply	612.85
	Hydroelectric	634.04
	Other	545.55
	Tailings	652.95
Lyndrassia	Debris Control	578.31
	Irrigation	611.33
	Recreation	597.39
	Flood Risk Reduction	584.32
	Fire Protection, Stock, Or Small Fish Pond	590.94
	Fish and Wildlife Pond	553.75
	Water Supply	612.73
	Hydroelectric	579.04
	Other	576.58
	Tailings	581.53
	Grade Stabilization	564.96
Navaldia	Debris Control	440.69
	Irrigation	516.27
	Recreation	539.85
	Flood Risk Reduction	506.62
	Fire Protection, Stock, Or Small Fish Pond	513.92
	Fish and Wildlife Pond	576.82
	Water Supply	526.06
	Hydroelectric	570.55
	Other	512.30
	Tailings	598.61
	Grade Stabilization	533.39

The amount of claim payout for the insurance company will be determined in the following way:

$$Expected\ Payout = \min(Loss * Percentage\ of\ Coverage, Coverage\ Limit)$$

2.3 Project Implementation Timeline

Note: Annual Inspect is represented by AI



The initial phase of the Veritas National Dam Insurance Program (VNDIP) will prioritize the rehabilitation of selected dams during the first two years. These dams, often older or in compromised conditions, are proven to be posing significant threats to public safety and economic stability. By addressing these high-risk structures early, the program seeks to reduce the potential for catastrophic incidents while establishing a foundation for broader infrastructure improvements (reduced economic loss, more available resources).

After the rehabilitation phase, the program will shift its focus to retrofitting tasks from the third to the fifth policy year. Retrofitting efforts will target dams that require upgrades to comply with current safety standards and to address potential environmental or operational challenges. Although these dams may not face immediate risk, proactive reinforcement will prevent future deterioration and failure. In the remaining years, the program will concentrate on the repair of selected dams, tackling less urgent maintenance needs. This phased approach ensures efficient resource allocation, prioritizing the most vulnerable dams within the portfolio.

According to the time length of different dam upgrading stages, we can measure the progress of the insurance with 2 evaluation periods. The short-term period, which covers the first 2 years, will emphasize the measurement of the initial project implementation efforts. The long-term period, which covers the subsequent years, will shift the focus to evaluating broader aspects.

1. **Short-Term Period (First 2 Years):**
    - **Enrollment Efficiency**
    - **Early Claims Processing:** Ensuring quick and accurate responses to dam incidents.
    - **Inspection & Rehabilitation Rollout:** Ensuring that high-risk dams are promptly assessed and improved.
2. **Long-Term Period (Subsequent Years):**
    - **Risk Reduction:** Measuring reduction in dam failure probability and the resulting economic loss.
    - **Infrastructure Stability:** Monitoring dam condition ratings and maintenance requirements.
    - **Financial Viability & Economic Resilience:** Tracking decreases in recovery costs and safeguarding economic activities tied to dams.

All the indicators align with **key performance metrics outlined in *Section 1.6***, and progress will be reviewed annually, allowing the program to achieve its goals over its entire duration while enabling adaptive management.

2.4 The Termination of Insurance Coverage

Following the full dam safety inspections conducted by our partnered engineering companies, we identify specific maintenance needs and underlying issues for each dam.

Regular maintenance is critical to prevent deterioration and extend a dam's service period. Based on the inspection findings, maintenance activities are classified into two categories:

Type	**Condition**
Immediate Maintenance	The dam is about to be overtopped or is already experiencing overtopping.
	Dams are nearing breach due to progressive erosion, slope failure, or other critical issues.
	Signs of internal erosion are present.
	Obstructed spillway.
	Signs of excessive seepage are present.
Required Maintenance	Overgrown underbrush and trees require removal, with grass cover being re-established or restored.
	Areas that have eroded need restoration.
	Defective spillways, gates, and inlet/outlet valves require repair or replacement.
	Metal components need proper maintenance.

Note: The table above only shows some of the commonly faced conditions. The actual classification of maintenance activities will be determined by the engineers from our partnered companies.

The insurance coverage will be terminated after **the event of dam failure and claims are provided**, or **the dam fails to comply with the maintenance schedule following a full inspection.**

Immediate Maintenance	Complete the required maintenance tasks and submit the necessary proof and reports to both the insurance company and the partnered engineering company within **3 months**.
Required Maintenance	Complete the required maintenance tasks and submit the necessary proof and reports to both the insurance company and the partnered engineering company within **1 year**.

Note: The time frame may be shortened based on the actual condition of the dam.

The maintenance schedule below details the key tasks required for dam upkeep (*please note that this list is not exhaustive*):

1. **Dam Bodywork Maintenance**
   - Riprap maintenance caused by repeated water strikes or weather-induced processes like freezing-thawing cycles.
   - Repair wave-induced surface erosion and seepage-induced erosion.
   - Seal and repair cracks.
   - Maintain and repair the surface and joint sealing systems.
   - Lubricate all moving parts.

2. **Appurtenant Structures Maintenance**
   - Remove debris and vegetation from spillway and outlet channels.
   - Repair damage such as cracking or erosion in spillways, stilling basins, and outlet structures.
   - Maintain booms to prevent debris from entering spillway outlets and water intake facilities.
   - Remove aquatic weed growth in the reservoir.
   - Apply appropriate coatings to extend the lifespan of existing piping systems.
   - Ensure the reservoir's inflow and outflow systems are fully functional.

3. **Drainage Systems**
   - Regularly clean and maintain the internal drainage systems at the dam's foundation, abutments, or other critical structures.

4. **Vegetation Control**
   - Remove unwanted vegetation and ensure any resulting holes are properly filled.
   - Fill any large holes that could cause erosion on the dam's body.
   - Prevent root penetration, as they may create pathways for water, leading to erosion.
   - Ensure that no vegetation will obstruct the visual inspection of the dam's condition.
   - Properly manage and dispose of plant cuttings.

5. **Maintenance of Dam Safety-Related Infrastructure:**
   - Repair access roads and remove excess roadside vegetation.
   - Maintain safety warning signs and fences surrounding the dam.

Section 3: Financial Result

3.1 The Determination of Premium

The premium for our financial product is calculated through a structured process that ensures all potential costs are covered while maintaining profitability. This involves two key steps: calculating the expected loss and determining the final premium. Each step incorporates critical financial factors to ensure fairness and sustainability.

1. Calculating the Actuarial Present Value of Expected Loss

The foundation of the premium is the actuarial present value of the expected loss, which represents the anticipated cost of claims or payouts. This figure is determined by considering several factors:

- **Expected Payout**: The amount that the insurer agrees to cover (*See section: 2.2.3*).
- **Annual Predicted Probability of Failure**: Since the probability is a 10-year probability, the following conversion is used to get the annual probability:

$$Annual\ Probability\ of\ Failure = 1 - (1 - Probability\ of\ Failure)^{0.1}$$

- **Inflation Factor**: $(1 + r)^t$ Based on rates from *Table 3*.
- **Discount Factor**: $(1 + i)^{-t}$ Based on the rates from *Table 3*.
- **Risk Factor**: This is an adjustment on the expected loss based on the hazard index classification, ensuring that the premium not only covers the expected loss but also provides a buffer for potential adverse outcomes.

Hazard Rating	Risk Factor
Low	1.0
Medium	1.1
High	1.2

To calculate the actuarial present value of expected loss for each dam, we multiply the factors at different times and then sum them up.

$$APV\ Expected\ Loss = \sum \begin{matrix} Expected\ Payout_t * Inflation\ Factor_t * Discount\ Factor_t * \\ Risk\ Factor_t * Annual\ Predicted\ Probability\ of\ Failure \end{matrix}$$

2. Determining the Final Premium

Once the actuarial present value of expected loss is calculated, we incorporate additional costs to get the final premium. This process involves the following components:

- **Actuarial Present Value of Inspection and Upgrade Cost**: The actuarial present value of all expenses related to dam inspection service and upgrade provided to the dams.
- **Pure Premium**: The base amount required to cover the expected loss and cost for providing dam inspection and upgrade.
- **10-Year Contingent Annuity Due**: Present value of 10 contingent annual premiums payable at the beginning of each year. The payment stops after the policy is terminated.

$$APV\ Expected\ Loss + APV\ Inspection\ and\ Upgrade\ Cost$$
$$= Pure\ Premium * 10\ year\ Contingent\ Annuity\ Due$$

- **Adjustment Loadings**: Factors that account for operational, reinsurance expense, or other cash outflow.

Factor	Value
Expense and Reinsurance Loading	0.05
Other Adjustment Loadings	If Necessary

$$Premium = Pure\ Premium * (1 + Expense\ and\ Reinsurance\ Loading)$$

This calculation ensures that the premium covers all costs, including expected losses, asset maintenance, and future payments to ensure business operations and profitability (*See Appendix: The Computation of Annual Premium*).

3.2 The Financial Projection

Since the portion of the premium received will be used to cover the losses incurred, therefore ensuring the amount received is sufficient to cover a massive claim is crucial. Since the frequency of dam failure is not available, we will use the concept of Monte Carlo Simulation here:

We first fix the annual probability of failure calculated from *Section 3.1* into a beta distribution, a continuous probability distribution defined on the interval [0,1] using the fitdist() in R. This function will find the best fit shape parameter 1 and shape parameter 2 for beta distribution; both parameters will be used to get a randomly generated probability using rbeta(). After that, the probability will be used to generate a binomial result, "Yes" or "No", using rbinom() to simulate the frequency of dam failure.

(*See Appendix: Monte Carlo Simulation for Dam Failure*)

The following assumptions will be used for financial projection:

No.	Assumption	Reason	Impact
1	The cost of dam upgrading services is evenly distributed across the designated period and paid at the beginning of each year.	This simplifies the cash flow calculation by avoiding interest between integer periods. We acknowledge that there may not be enough engineering companies in Tarrodan to upgrade and inspect all dams promptly.	Small impact, since the APV of the cost to upgrade each dam is minor compared to the APV of expected loss per dam.
2	The cost of dam inspection services is evenly distributed across the designated period and paid for at the beginning of each year.	This assumption is made to simplify cash flow calculations.	Small impact.

Table 15: Financial Projection

Year	Premium Received (in Qm)	Claims Incurred (in Qm)	Expense Incurred (in Qm) *	Gross Profit (in Qm) **
1	40,419.63	37,928.24	2,047.50	443.89
2	40,067.04	34,899.25	2,045.90	3,121.89
3	39,731.63	33,178.39	2,548.90	4,004.34
4	39,392.19	30,038.67	2,547.50	6,806.02
5	39,096.43	26,805.00	2,546.50	9,744.93
6	38,855.87	29,856.83	512.87	8,486.17
7	38,588.85	36,176.54	511.27	1,901.04
8	38,236.14	34,083.93	509.87	3,642.34
9	37,896.32	34,290.40	508.27	3,097.65
10	37,574.64	27,951.39	506.97	9,116.28

Note (*): The Expense Incurred includes the cost to inspect and upgrade the dams.

Note (**): The Gross Profit = Premium Received - Claims Incurred - Expenses Incurred

Figure 10: Financial Projection (Visualized)

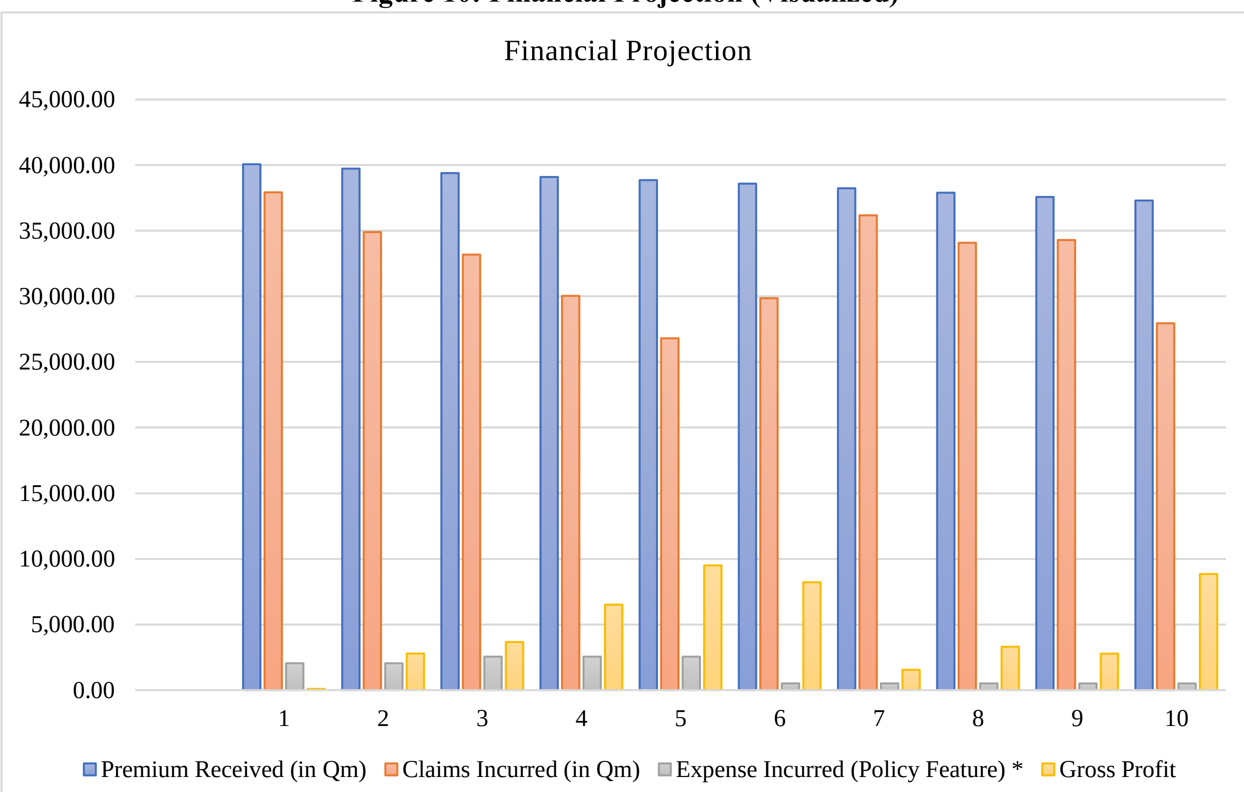


Table 16: First Year Premium Distribution by Region and Primary Purpose

Region	Primary Purpose	Sum of Annual Premium (in Qm)	Weightage (% of Total Premium Received)
Flumevale	Debris Control	133.90	0.3313
	Fire Protection, Stock, Or Small Fish Pond	216.79	0.5363
	Fish and Wildlife Pond	286.55	0.7089
	Flood Risk Reduction	943.84	2.3351
	Hydroelectric	363.05	0.8982
	Irrigation	2672.98	6.6131
	Other	524.90	1.2986
	Recreation	595.80	1.4740
	Tailings	28.54	0.0706
	Water Supply	3102.11	7.6748
Lyndrassia	Debris Control	119.59	0.2959
	Fire Protection, Stock, Or Small Fish Pond	5415.93	13.3993
	Fish and Wildlife Pond	307.96	0.7619
	Flood Risk Reduction	1762.17	4.3597
	Grade Stabilization	851.17	2.1058
	Hydroelectric	5.88	0.0146
	Irrigation	2552.93	6.3161
	Other	453.65	1.1224
	Recreation	2427.71	6.0063
	Tailings	81.88	0.2026
	Water Supply	263.37	0.6516
Navaldia	Debris Control	29.66	0.0734
	Fire Protection, Stock, Or Small Fish Pond	755.73	1.8697
	Fish and Wildlife Pond	69.23	0.1713
	Flood Risk Reduction	6911.52	17.0994
	Grade Stabilization	47.62	0.1178
	Hydroelectric	27.61	0.0683
	Irrigation	1415.94	3.5031
	Other	811.38	2.0074
	Recreation	5095.06	12.6054
	Tailings	137.51	0.3402
	Water Supply	2007.68	4.9671

Table 17: APV of Inspection and Upgrade Cost Distribution by Region and Primary Purpose

Region	Primary Purpose	Sum of APV (in Qm)	Weightage (% of Total APV)
Flumevale	Debris Control	14.08	0.1039
	Fire Protection, Stock, Or Small Fish Pond	107.67	0.7941
	Fish and Wildlife Pond	79.93	0.5895
	Flood Risk Reduction	219.04	1.6155
	Hydroelectric	65.05	0.4798
	Irrigation	841.12	6.2035
	Other	65.39	0.4823
	Recreation	187.58	1.3834
	Tailings	1.52	0.0112
	Water Supply	374.79	2.7642
Lyndrassia	Debris Control	37.11	0.2737
	Fire Protection, Stock, Or Small Fish Pond	2357.69	17.3887
	Fish and Wildlife Pond	141.85	1.0462
	Flood Risk Reduction	384.53	2.8361
	Grade Stabilization	257.97	1.9026
	Hydroelectric	0.34	0.0025
	Irrigation	840.32	6.1976
	Other	121.34	0.8949
	Recreation	530.48	3.9125
	Tailings	6.04	0.0445
	Water Supply	46.41	0.3423
Navaldia	Debris Control	9.89	0.0729
	Fire Protection, Stock, Or Small Fish Pond	395.54	2.9173
	Fish and Wildlife Pond	17.64	0.1301
	Flood Risk Reduction	2800.75	20.6563
	Grade Stabilization	14.01	0.1034
	Hydroelectric	4.55	0.0335
	Irrigation	599.57	4.4220
	Other	209.63	1.5461
	Recreation	1916.37	14.1338
	Tailings	28.73	0.2119
	Water Supply	881.82	6.5037

Table 18: First Year Premium Distribution by Hazard and Dam Condition Assessment

Hazard	Assessment	Sum of Annual Premium (in Qm)	Weightage (% of Total Premium Received)
High	Fair	8156.44	20.1794
	Poor	3230.09	7.9914
	Satisfactory	8558.40	21.1739
	Unsatisfactory	3942.72	9.7545
Low	Fair	1744.79	4.3167
	Poor	1014.76	2.5106
	Satisfactory	2212.14	5.4729
	Unsatisfactory	2037.81	5.0416
Significant	Fair	2161.38	5.3474
	Poor	945.49	2.3392
	Satisfactory	4607.68	11.3996
	Unsatisfactory	1807.92	4.4729

Table 19: First Year Premium Distribution by Region and Age

Region	Age	Sum of Annual Premium (in Qm)	Weightage (% of Total Premium Received)
Flumevale	0-50	1690.38	4.1821
	50-100	4617.81	11.4247
	100-150	2437.48	6.0304
	150-200	122.78	0.3038
Lyndrassia	0-50	4098.75	10.1405
	50-100	9694.04	23.9835
	100-150	436.19	1.0791
	150-200	5.94	0.0147
	200-250	7.33	0.0181
Navaldia	0-50	4276.41	10.5800
	50-100	11798.93	29.1911
	100-150	1151.04	2.8477
	150-200	75.06	0.1857
	200-250	6.90	0.0171
	250-300	0.59	0.0015

Table 20: APV of Inspection and Upgrade Cost Distribution by Hazard and Dam Condition Assessment

Hazard	Assessment	Sum of APV (in Qm)	Weightage (% of Total APV)
High	Fair	115.00	0.8481
	Poor	1220.08	8.9984
	Satisfactory	612.35	4.5163
	Unsatisfactory	8.76	0.0646
Low	Fair	695.52	5.1297
	Poor	3848.26	28.3821
	Satisfactory	172.65	1.2733
	Unsatisfactory	1.72	0.0127
Significant	Fair	5.93	0.0438
	Poor	569.33	4.1990
	Satisfactory	5916.02	43.6324
	Unsatisfactory	363.55	2.6813

Table 21: APV of Inspection and Upgrade Cost Distribution by Region and Age

Region	Age	Sum of APV (in Qm)	Weightage (% of Total APV)
Flumevale	0-50	2466.98	18.1947
	50-100	1263.84	9.3212
	100-150	647.23	4.7735
	150-200	1571.17	11.5878
Lyndrassia	0-50	1319.81	9.7340
	50-100	1042.93	7.6919
	100-150	66.55	0.4908
	150-200	1482.96	10.9373
	200-250	1114.06	8.2165
Navaldia	0-50	1021.10	7.5309
	50-100	133.50	0.9846
	100-150	1428.66	10.5368
	150-200	2466.98	18.1947
	200-250	1263.84	9.3212
	250-300	647.23	4.7735

Regional Funding Strategies for the Veritas National Dam Insurance Program

VNDIP is intended to offer premium deductions to dam owners that follow the policy of this program through external funding, making this program more appealing. The external funding will be sourced through a combination of taxes and government subsidies. This approach requires collaboration between the insurance company and legislative authorities to ensure effective execution. The regional funding strategy should be tailored based on the economic composition as well as the geographical conditions of each region. The following table provides a brief overview of the Tarrodan's economic situation in 2023:

Distribution of Gross Domestic Product (GDP) by Industry (2023)

Industry	Flumevale	Lyndrassia	Navaldia
Agriculture	5.90%	1.90%	0.70%
Entertainment	4.70%	4.60%	3.60%
Construction	4.10%	4.90%	5.90%
Education	7.60%	10.10%	9.00%
Finance	18.80%	15.50%	18.80%
Information	7.50%	3.20%	3.90%
Manufacturing	10.00%	10.90%	11.30%
Mining	0.70%	8.80%	0.70%
Professional Services	12.20%	10.20%	12.90%
Retail	6.00%	6.90%	6.20%
Transportation	3.30%	3.70%	5.50%
Utilities	1.40%	1.50%	1.80%
Trade	5.40%	6.70%	8.80%
Government	10.60%	8.70%	9.00%
Other	1.80%	2.40%	1.90%

Population Overview

Year	Flumevale	Lyndrassia	Navaldia	Tarrodan
2019	45,363,514	7,067,855	39,808,697	92,240,066
2020	45,502,051	7,097,789	40,175,188	92,775,028
2021	45,651,175	7,131,024	40,565,887	93,348,086
2022	45,599,000	7,157,446	40,953,108	93,709,554
2023	45,311,937	7,239,138	42,148,205	94,699,280

Population Distribution

Category	Flumevale	Lyndrassia	Navaldia
Urban (>50,000)	88.20%	52.40%	38.80%
Urban (<50,000)	5.70%	16.00%	23.90%
Rural	6.10%	31.60%	37.30%

Distribution of Number of Housing Units by Value

Value	Flumevale	Lyndrassia	Navaldia	Tarrodan
Less than Q50,000	446,720	206,960	871,223	1,524,903
Q50,000 to Q99,999	552,111	414,955	1,593,970	2,561,036
Q100,000 to Q149,999	887,815	471,356	2,169,919	3,529,090
Q150,000 to Q199,999	1,219,671	448,956	2,505,863	4,174,490
Q200,000 to Q299,999	2,730,435	572,280	3,816,207	7,118,922
Q300,000 to Q499,999	2,521,934	532,740	3,189,802	6,244,476
Q500,000 to Q999,999	5,165,736	268,786	1,825,774	7,260,296
Q1,000,000 or more	2,386,698	51,349	351,150	2,789,197
Total	15,911,120	2,967,382	16,323,908	35,202,410

I. Property Tax in Disaster-Prone Zones

Charge the construction company 1% of the value of new developments in region-specific high-risk zones:

- **Flumevale**: Riverside constructions in flood-prone areas.
- **Lyndrassia**: Constructions in avalanche- and earthquake-prone zones, such as foothills and unstable mountain slopes.
- **Navaldia**: Constructions in coastal areas vulnerable to tropical storms and tsunamis that result in flooding.

From the *Distribution of Number of Housing Units by Value table*, many houses in Flumevale and Navaldia exceed the **Q300,000+ categories**. We can expect that this tax will generate substantial funding.

This idea is inspired by Vollebergh and Djik (2016), which discuss how taxes are applied to both residential and commercial properties in the Netherlands to finance flood protection and water management measures.

II. Usage-Based Agricultural Irrigation Tax

This strategy is suitable for Flumevale. Due to its expansive forests and fertile plains, 5.9% of its GDP comes from agricultural activities. We can expect that the water demand for agricultural activities is significant in this region. Flumevale could introduce a tiered water usage tax for companies and individual farmers to fund the national insurance program. For example, 0.05 Q per cubic meter for the first 10,000 m^3 and 0.15 Q per cubic meter for extra usage. This approach not only funds the program but also indirectly encourages the use of water-saving practices like drip irrigation.

III. Usage-Based Urban Water Supply Tax

Inspired by Singapore's Water Conservation Tax, the nation could implement a tax on urban households based on their water consumption. This approach ensures stable funding for the insurance program while reducing the per capita water use to conserve the environment.

IV. Tourism Activity Tax

Referring to *Table 5: Composition of Earthen Dams by Region and Purpose*, many dams are used for recreational purposes, with 224 dams in Flumevale, 1983 dams in Lyndrassia, and 2795 dams in Navaldia. Taxes can be imposed on recreational services provided at these dams, such as boat rentals (the Flumevale River Racing) and guided tours.

Section 4: Assumptions - Detailed Explanation

A series of assumptions are used throughout the design of the policy features and financial projection for the Veritas National Dam Insurance Program (VNDIP).

These assumptions affect the estimated costs, the determination of premiums, and the measurement of the program's success. In this section, each major assumption will be explained in detail, ensuring a clear understanding of how they affect the analysis.

No.	Assumptions	Description	Solution
1	Constant Annual Probability of Failure	Our team calculates the annual failure probability using the formula from *Section 3.1 The Determination of Premium*, which simplifies the complex risk model into a constant value for projections. However, this method ignores time-related factors such as gradual wear and environmental degradation. A dam could have a lower risk in its early years, but the risk gets significantly higher due to cumulative stress. This assumption will not reflect this situation. If these risks are underestimated in later years, the company could face unexpectedly high claims.	This is one of the main reasons the program lasts only 10 years. Since most dams are designed to operate for 50 to 100 years, the insurance program can be restructured or readjusted over time to account for the effects of aging and evolving risks.
2	Standardized Inspection Cost	The costs are taken from external benchmarks. This assumption assumes standardized inspection cost for all dams with the same hazard classification (*See Section 2.2.1*). However, inspection costs can vary due to differences in dam height, purpose, volume, and even the region, considering factors like equipment availability and the geographical challenges of inspecting remote sites. If Tarrodan's actual costs exceed these benchmarks, we risk underestimating both our profit and premiums, potentially leading to unexpected losses.	To tackle this, our team will gather quotations from our partner engineering firms and perform detailed cost analysis. We will then adjust the premium amounts based on the latest cost.

3	Standardized Upgrade Cost	The financial projection for dam upgrade costs is based on categories defined by dam height and condition (*See Section 2.2.2*). These estimated costs are also taken from external benchmarks. While this assumption simplifies the financial projection, it may not fully capture the complexities of real-world scenarios. For instance, upgrading costs can vary if a dam has unique issues such as foundation instability, environmental challenges, or geographical factors like material availability. Additionally, different regions in Tarrodan face distinct risks; for example, dams in Flumevale may deal with periodic flooding, whereas those in Lyndrassia could be more prone to avalanches, leading to different upgrade costs.	To address these variations, we require a detailed dam condition assessment report before any dam joins the insurance program. These reports will be provided to our partner engineering companies to determine the actual upgrade costs.
4	Inflation and Discount Rates Follow Historical Trends	Forecasted rates are simulated from the Cox Ross Ingersoll (CIR) Model, which estimates interest rate movements using fine-tuned parameters from historical data (*See Section 1.4*). However, inflation and interest rates are subject to volatility from external factors such as policy shifts, market disruptions, or crashes. The impact of this assumption can be summarized as follows: I. If the inflation rate turns out higher than the projected rates, the actuarial present values may be understated, leading to insufficient premiums. II. If the inflation rate is lower than the projected rates, the actuarial present	Our team will review Tarrodan's economic situation annually and update our financial projections with the latest rates.

		values may be overstated, leading to excessive premiums. III. If the interest rate is higher than the projected rates, the actuarial present values may be overstated, leading to excessive premiums. IV. If the interest rate is lower than the projected rates, the actuarial present values may be understated, leading to insufficient premiums.	
5	Constant Risk Factors	Risk adjustments are applied using factors of 1.0 for Low, 1.1 for Medium, and 1.2 for High hazard ratings to scale expected losses and premiums according to dam risk levels. This assumption simplifies the computational process. However, this assumption may not be able to fully capture the exponential risk escalation for dams due to compounding factors. The underestimation of these risks often leads to underpricing, leaving the program underfunded for major claims.	Our team will address this issue by introducing the Hazard Index and quantile-based underwriting criteria to prevent adverse selection. The insurance program also mandates the evaluation of the dams' actual condition before they can enroll in the insurance program.

Section 5: Risks and Risk Mitigation Considerations

5.1 Risk Assessment

The Veritas National Dam Insurance Program (VNDIP) faces risks from technical, operational, financial, and environmental aspects. The table below is a brief analysis of the key risks, along with their potential impacts, likelihoods, and risk mitigation strategies.

Before that it is important to recognize the difference between quantifiable and qualitative risk:

- Quantifiable risks are the risks that can be measured numerically using historical records, statistical models, or other metrics.
- Qualitative risks are the risks that are not easily quantified due to lack of data or because the risks are subjective. Qualitative risk analysis relies on expert judgment and subjective criteria.

No.	Risk	Type	Impact (1-5)	Likelihood (1-5)	Description	Mitigations
1	Predictive Model Inaccuracy	Quantifiable	5	3	The XGBoost model used for predictions may underestimate the probability and the loss amount due to shifting external factors.	Update the model regularly with the real-world failure data. Ensure the outputs are validated or consider using a more advanced algorithm.
2	Climate Change Impacts	Quantifiable	5	2	The increasing frequency of floods or droughts due to changes in climate factors (rainfall, snowfall, average temperature) may cause the dam to no longer meet local demands.	Conduct regular inspections, maintenance, and upgrades on the dams. Incorporate an advanced model that accounts for environmental changes.
3	Reputational Damage	Qualitative	5	1	High-profile dam failures under the coverage may cause the public to question the effectiveness of the insurance program	Publish annual insurance reports and investigation reports after dam incidents.

					and eventually lead to a loss of trust.	
4	Data Imputation Errors	Quantifiable	4	4	The imputed data may be incorrect, thereby affecting the model's accuracy and resulting in incorrect analysis.	Perform sensitivity analyses on the imputed values.
5	Inflation or Interest Rate Shocks	Quantifiable	4	3	Sudden changes in external market factors may cause the projected rates to be incorrect.	Use combinations of derivative securities such as interest rate swaps to mitigate the risk of inflation and interest rate fluctuations. Regularly calibrate the model and adjust premiums.
6	Misjudgment by Engineers	Qualitative	4	2	Errors in dam condition assessments or inspection reports. For example, the overlooking of cracks and seepage can lead to inadequate repair and delayed remedial actions.	Implement strict and peer-reviewed protocols for inspections. Validate assessments using professional tools. Mandate regular professional training for engineers and operational personnel.
7	Fraudulent Practices	Qualitative	4	1	Engineers or underwriting personnel may accept bribery to falsify dam condition assessments before and after joining the insurance program.	Cross-validate inspection reports. Establish whistleblower protection and reward within the company.

8	Engineering Capacity Shortages	Qualitative	3	3	Delays in dam inspections or upgrades due to the shortage of qualified engineers.	Partnered with international engineering companies. Propose incentive-based training programs for dam engineers to the government.
9	Non-Compliance by Dam Owners	Qualitative	3	2	Mandatory maintenance may be neglected by dam authorities.	Coverage Termination (*See: Section 2.4*).
10	Reinsurance Market Volatility	Quantifiable	2	4	A sudden increase in reinsurance costs could lead to insufficient funding of the program.	Diversify this risk by cooperating with different reinsurers across multiple regions.

Risk Map

		Impact				
		1	2	3	4	5
Likelihood	1				Fraudulent Practices	Reputational Damage
	2			Non-Compliance by Dam Owners	Misjudgment by Engineers	Climate Change Impacts
	3			Engineering Capacity Shortages	Inflation or Interest Rate Shocks	Predictive Model Inaccuracy
	4		Reinsurance Market Volatility		Data Imputation Errors	
	5					

5.2 Sensitivity Analysis

Figure 11: The coverage limit threshold increases from the 70th percentile to the 75th percentile.

Result: The total annual premium received increased by 1.87%.

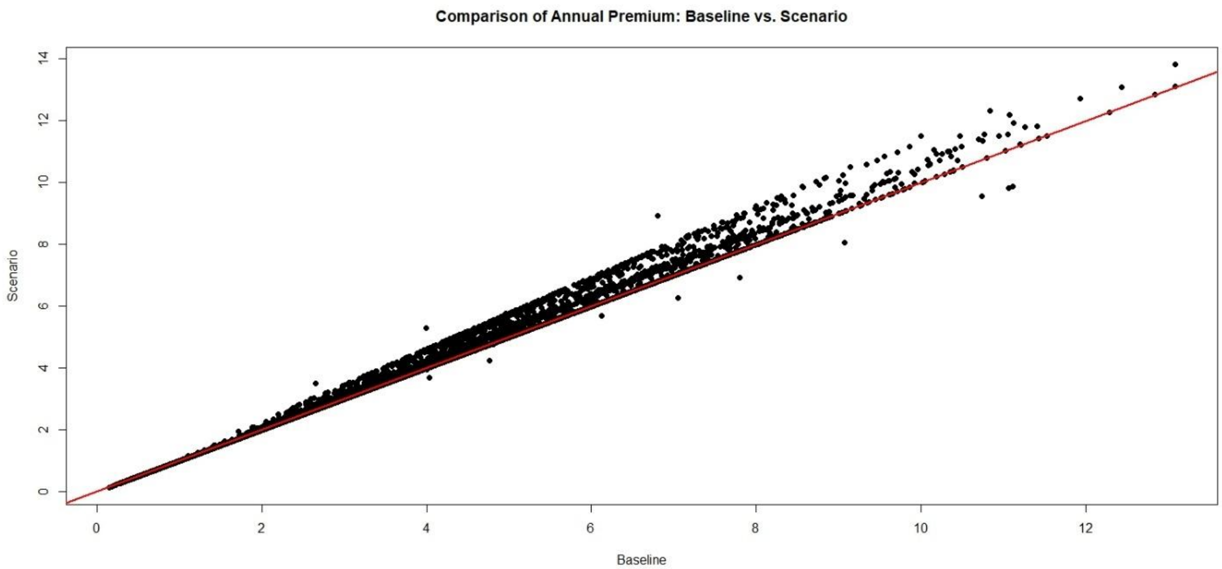


Figure 11

Note: The red line represents the baseline (the original situation). The distance between the dots and the red line indicates the dispersion of the scenario from the baseline.

We can clearly see that as the values (annual premium, in Qm) on the X-axis are getting closer and closer to the threshold, the dispersion is more and more obvious.

Figure 12: The coverage limit threshold decreases from the 70th percentile to the 65th percentile.

Result: The total annual premium received decreased by 2.365%.

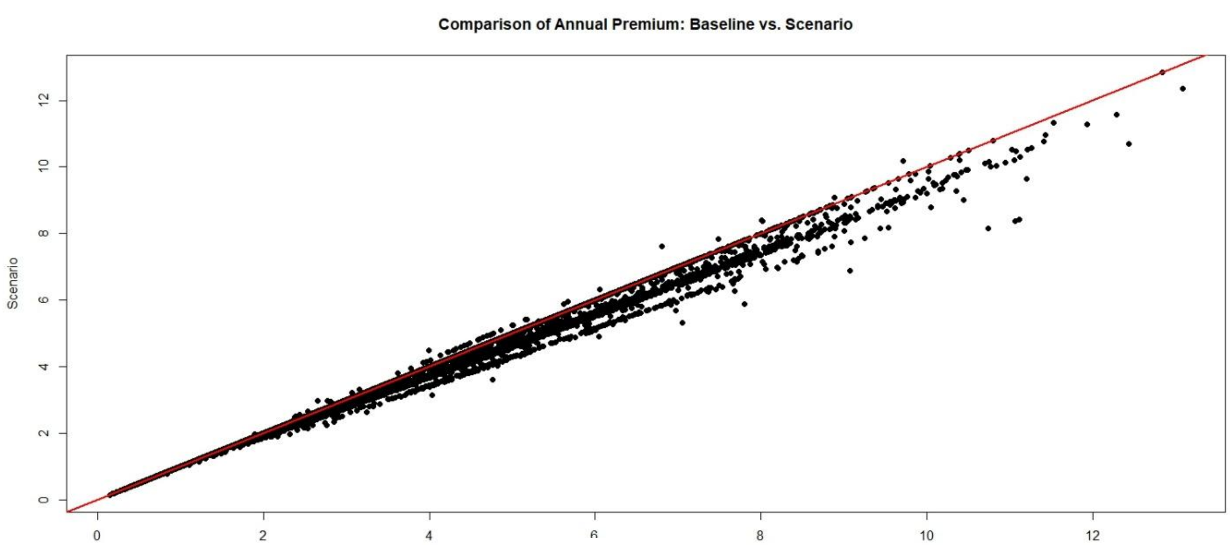


Figure 12

Figure 13: The 10-year probability of failure for each dam increased by 1%.

Result: The total annual premium received increased by 13.33%.

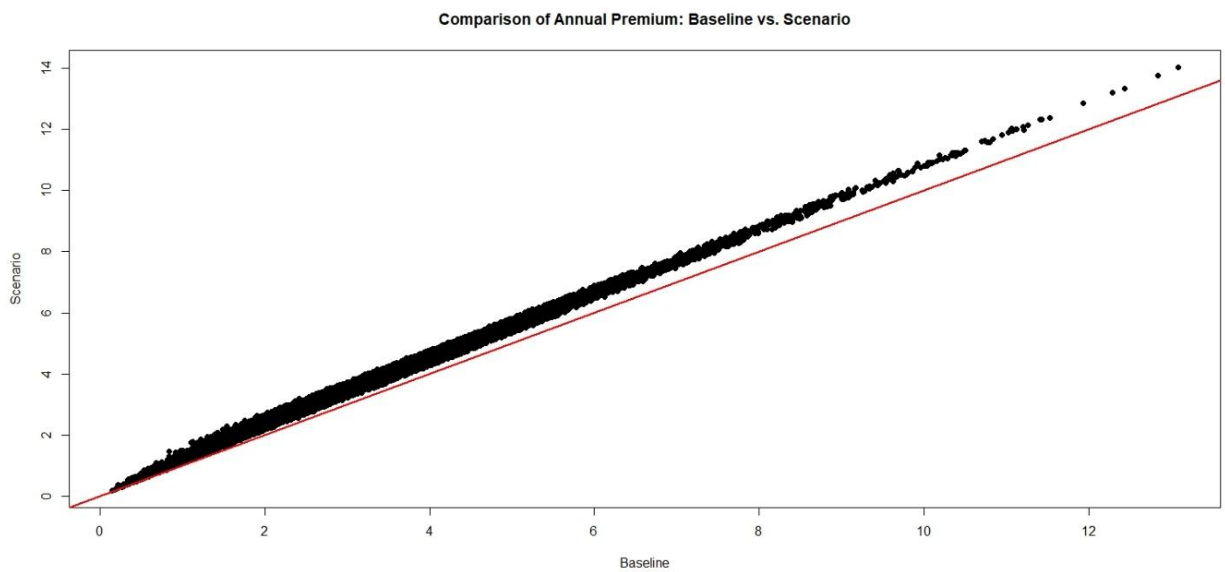


Figure 13

Figure 14: The 10-year probability of failure for each dam decreased by 1%.

Result: The total annual premium received decreased by 13.09%.

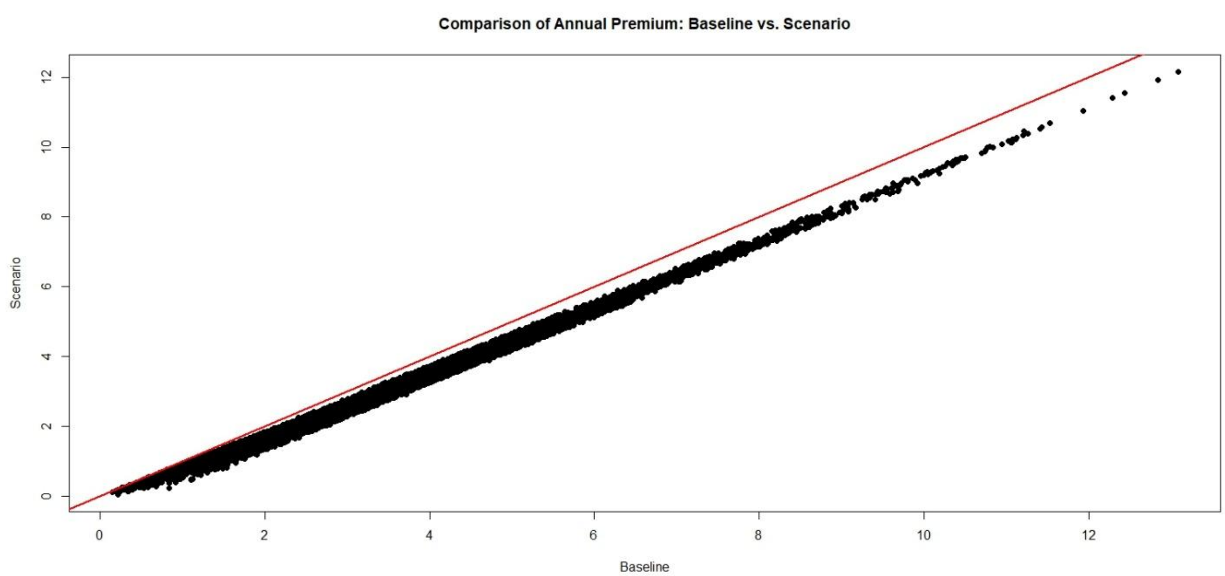


Figure 14

Figures 13 and 14 highlight the importance of predictive model accuracy. Even a 1% shift in the 10-year probability of failure can lead to significant overcharging or underfunding of the insurance program.

Figure 15: The inflation rate for each year increased by 5%.

Result: The total annual premium received increased by 4.583% as a result of the increase in the APV of inspection cost (+21.8437%), the increase in APV of upgrade cost (+7.561%), and the increase in the APV of expected loss (+24.522%).

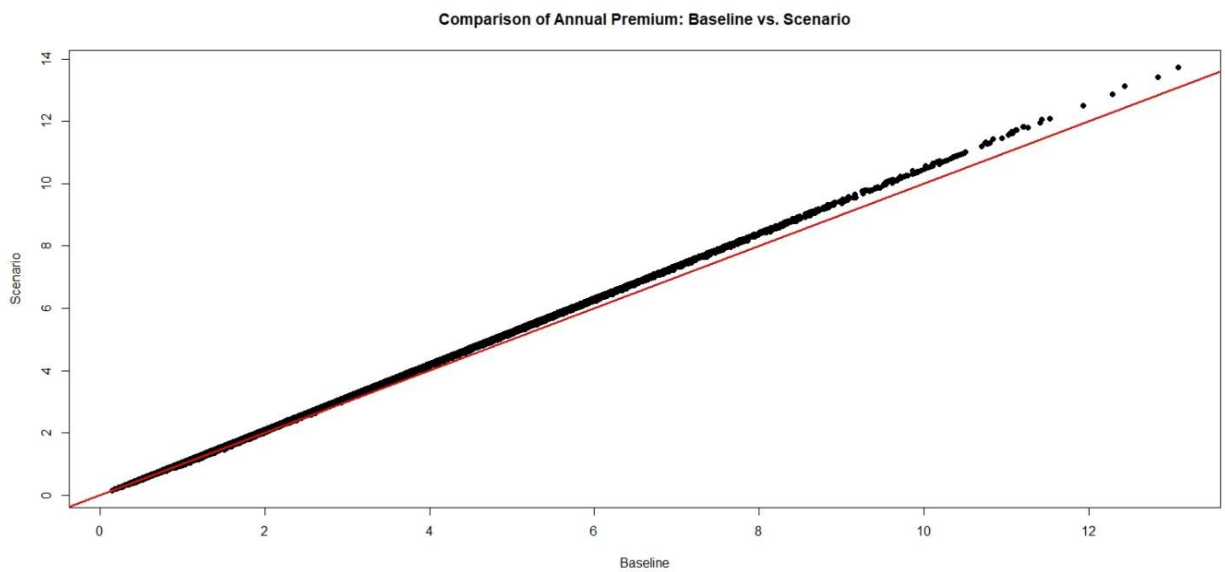


Figure 15

Figure 16: All the costs of inspection and upgrade increased by 20%.

Result: The total annual premium received increased by 0.812%.

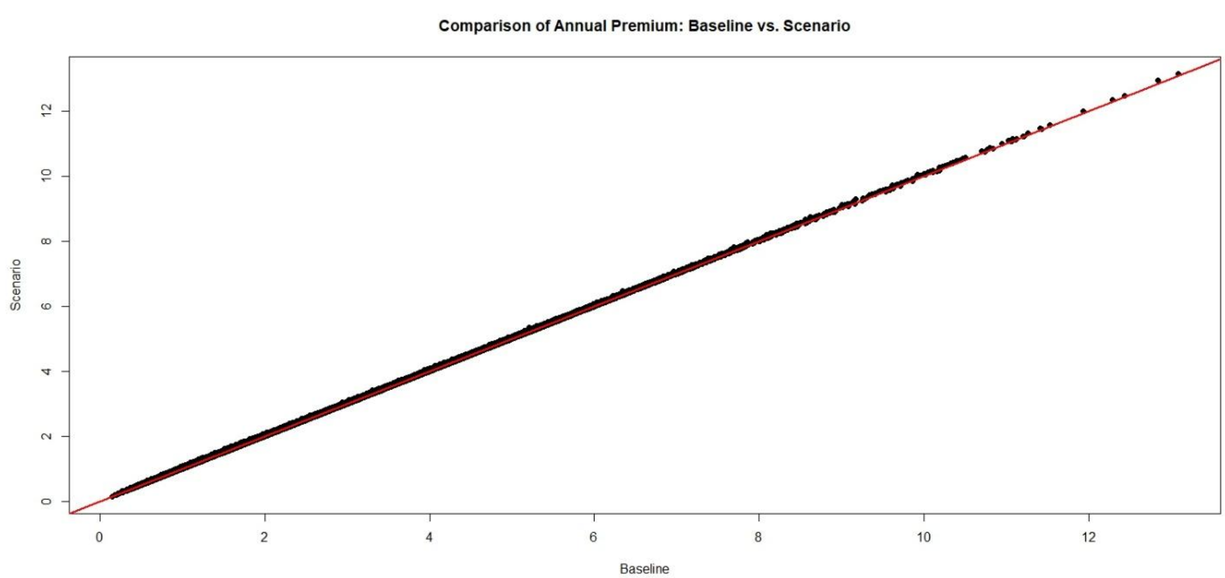


Figure 16

Climate Change Risk Considerations

Incorporating climate change factors into the design of the Veritas National Dam Insurance Program (VNDIP) is a significant challenge due to the incompleteness of climate-related data for each region, such as the changes in rainfall patterns and rising sea levels that could bring serious impacts to the dam's safety. Despite facing this challenge, the program will ensure that all insured dams are protected while complying with the national environmental regulations (*See Section 2.1 Underwriting Criteria and Coverage Eligibility for Dam Insurance*).

VNDIP will ensure that all dam upgrades are conducted in ways that minimize the environmental impact; all the scraps, debris, and outdated equipment will be handled according to the national environmental regulations.

Section 6: Data Limitations

The analysis and statistics supporting the measurement and effectiveness of the Veritas National Dam Insurance Program (VNDIP) are affected by several data limitations. These limitations will affect the reliability of financial projections. We conclude that these limitations arise from 3 major issues:

1. Incomplete data.
2. Reliance on the historical trends.
3. External benchmarks that are not tailored to Tarrodan's context.

No.	Data Limitation	Description	Solution
1	Missing Data and Imputation	From *Section 1.2 Method and Assumptions on Data Imputations*, it is clear the dataset has significant gaps in the essential information needed to evaluate a dam. For instance, "Assessment" is missing in 12.19% of entries, "Last Inspection Date" in 48.18%, and "Assessment Date" in 46.97%. To handle these missing values, our team used Multivariate Imputation by Chained Equations (MICE), which predicts unknowns based on observed data patterns. However, MICE assumes data is missing at random, so imputed values may not always reflect the actual situation. For example, a dam with high loss given failure but no assessment data; logically, the missing value is likely to be "Poor" or "Unsatisfactory", but MICE may impute an overly optimistic result such as "Satisfactory", underestimating its true risk.	To address this, we adopt a conservative approach throughout the data imputation process, assigning the worst-case value for missing critical data. For example, any dam without an assessment rating is labeled as "Not Available." Besides, the risk factor is incorporated as a buffer in loss reserves.

2	Assumptions in Predictive Model	Our team uses the XGBoost model to predict the probability of failure and total loss given failure. This model assumes the relationships between key factors such as the height and the volume of the dam will remain the same over time to ensure accurate predictions. However, real-world changes like increased rainfall can alter these relationships. For example, in areas with heavier rain, dam height becomes even more vital in preventing overtopping, and any of these miscalculations may lead to underestimated losses. Our team also assumes that dam inspections and upgrades will lead to changes in hazard classification and assessment ratings.	To tackle this, our team will update the model regularly with the latest records and dam failure data. The predictions will also be validated through comparisons across different models.
3	External Cost Benchmarks	Estimated costs of dam inspections and upgrades are taken from external sources. However, local factors like labor costs, material prices, or regulatory requirements can differ.	Our team will conduct an additional cost study before implementing the insurance program.
4	Dam Failure Simulation Based on Historical Rates	Our team simulates dam failures by fitting the historical probability of failure into a probability distribution; this approach provides a solid statistical foundation. However, this method does not capture new factors like climate change, which can lead to more frequent flooding and add stress to dams, thereby increasing the chances of dam failure.	Our team plans to incorporate real dam failure records when available or use other advanced statistical methods for future projections.

5	Geographical and Regional Variability	The dataset consists of dams across 3 regions: Flumevale, Lyndrassia, and Navaldia. However, the geographical and regional differences are not available, like the rainfall, earthquake records, sea levels, and the location of the dams. Treating all dams uniformly could misrepresent the potential risk, therefore resulting in the mispricing.	Our team plans to obtain localized geographical condition studies and make necessary adjustments before the insurance program is implemented.

Appendix: Data Imputation

```
library(mice)

library(tidyverse)

library(readr)


dams <- readr::read_csv("C:/Users/userc/Desktop/Dam Data.csv")


# Replace all empty cells in Assessment and Spillway with the worst possible value

dams <- dams %>%

mutate(

   Assessment = replace_na(Assessment, "Not Available"),

   Spillway = replace_na(Spillway, "Uncontrolled")

  )


# For missing “Primary Purpose”, “Primary Type” we impute them with the most common value after
grouping to maintain the bias among the original dataset

dams <- dams %>%

  group_by(Region, Hazard, `Regulated Dam`) %>%

  mutate(`Primary Purpose` = replace_na(`Primary Purpose`,

                                        names(sort(table(`Primary Purpose`), decreasing = TRUE))[1])) %>%

  ungroup()


dams <- dams %>%

  group_by(`Primary Purpose`, Region, Hazard, `Regulated Dam`) %>%

  mutate(`Primary Type` = replace_na(`Primary Type`,

                                     names(sort(table(`Primary Type`), decreasing = TRUE))[1])) %>%

  ungroup()
```

```
# After replacing values of 0, we impute the missing height by fitting the medium height after grouping
dams <- dams %>%
  mutate(`Height (m)` = ifelse(`Height (m)` == 0, NA, `Height (m)`)) %>%
  group_by(`Primary Type`, `Primary Purpose`, Region, Hazard, `Regulated Dam`) %>%
  mutate(`Height (m)` = replace_na(`Height (m)`,
                                   median(`Height (m)`, na.rm = TRUE))) %>%
  ungroup()

# Check whether the columns above are fully imputed
colMeans(is.na(dams)) * 100

# Convert some columns/ variables into factor before data imputation
dams$Region <- as.factor(dams$Region)
dams$'Regulated Dam' <- as.factor(dams$'Regulated Dam')
dams$'Primary Purpose' <- as.factor(dams$'Primary Purpose')
dams$'Primary Type' <- as.factor(dams$'Primary Type')
dams$Hazard <- as.factor(dams$Hazard)
dams$Assessment <- as.factor(dams$Assessment)
dams$Spillway <- as.factor(dams$Spillway)

# Two Stage MICE Imputation
# Stage 1
# Define variables that need to be imputed and the variables that are used for prediction.
vars_1 <- c("Year Completed", "Primary Purpose", "Primary Type", "Height (m)", "Hazard",
"Assessment", "Probability of Failure", "Distance to Nearest City (km)", "Loss given failure - prop
(Qm)", "Loss given failure - liab (Qm)", "Loss given failure - BI (Qm)"
)

# After defining the variables we needed (not all columns are used, for example, the ID is not used), we
construct a data frame that only consists of the variables we needed
dams_subset_1 <- dams[, vars_1]
```

```
# Define the methods that will be used to impute the column (pmm means predictive mean matching)
meth_1 <- make.method(dams_subset_1)
meth_1["Year Completed"]  <- "pmm"
meth_1["Distance to Nearest City (km)"]  <- "pmm"
meth_1["Loss given failure - prop (Qm)"] <- "pmm"
meth_1["Loss given failure - liab (Qm)"] <- "pmm"
meth_1["Loss given failure - BI (Qm)"] <- "pmm"

# To avoid singularity, this function can pick the needed variables that have at least correlation of 0.05
pred_1 <- quickpred(dams_subset_1, mincor = 0.05)

# Perform Imputation
dams_imputed_stage1 <- mice(dams_subset_1, m = 5, maxit = 5, method = meth_1, predictorMatrix =
pred_1, seed = 1)

# Store the completed data into another data frame and write the results
dams_complete_1 <- complete(dams_imputed_stage1, 1)

dams$"Year Completed" <- dams_complete_1$"Year Completed"
dams$"Distance to Nearest City (km)" <- dams_complete_1$"Distance to Nearest City (km)"
dams$"Loss given failure - prop (Qm)" <- dams_complete_1$"Loss given failure - prop (Qm)"
dams$"Loss given failure - liab (Qm)" <- dams_complete_1$"Loss given failure - liab (Qm)"
dams$"Loss given failure - BI (Qm)" <- dams_complete_1$"Loss given failure - BI (Qm)"

# Stage 2
# Define variables that need to be imputed and the variables that are used for prediction.
vars_2 <- c("Year Completed", "Primary Purpose", "Primary Type", "Height (m)", "Hazard",
"Assessment", "Probability of Failure", "Distance to Nearest City (km)", "Loss given failure - prop
(Qm)", "Loss given failure - liab (Qm)", "Loss given failure - BI (Qm)", "Length (km) ", "Volume (m3) ",
"Surface (km2)", "Drainage (km2)"
)
```

After defining the variables we needed (not all columns are used, for example, the ID is not used), we construct a data frame that only consists of the variables we needed

dams_subset_2 <- dams[, vars_2]

Define the methods that will be used to impute the column (pmm means predictive mean matching)

meth_2 <- make.method(dams_subset_2)

meth_2["Length (km)"] <- "pmm"

meth_2["Volume (m3)"] <- "pmm"

meth_2["Surface (km2)"] <- "pmm"

meth_2["Drainage (km2)"] <- "pmm"

To avoid singularity, this function can pick the needed variables that have at least correlation of 0.05

pred_2 <- quickpred(dams_subset_2, mincor = 0.05)

Perform Imputation

dams_imputed_stage2 <- mice(dams_subset_2, m = 5, maxit = 5, method = meth_2, predictorMatrix = pred_2, seed = 2)

Store the completed data into another data frame and write the results

dams_complete_2 <- complete(dams_imputed_stage2, 1)

dams$"Year Completed" <- dams_complete_2$"Year Completed"

dams$"Length (km)" <- dams_complete_2$"Length (km)"

dams$"Volume (m3)" <- dams_complete_2$"Volume (m3)"

dams$"Surface (km2)" <- dams_complete_2$"Surface (km2)"

dams$"Drainage (km2)" <- dams_complete_2$"Drainage (km2)"

Export the result

write_csv(dams, "C:/Users/userc/Desktop/DamData_Processed.csv")

Appendix: Gradient Boost Predictive Model (XGBoost)

Predictive XGBoost Model performance:

Name	R-squared
xgb_model	0.48
xgb_model_loss	0.72

Model 1: xgb_model, used for probability prediction

```
library(xgboost)

library(Matrix)

library(SHAPforxgboost)


dams <- readr::read_csv("C:/Users/userc/Desktop/DamData_Processed2.csv")

dams <- dams %>%

mutate(

Hazard = factor(Hazard),  Assessment = factor(Assessment), `Regulated Dam` = factor(`Regulated Dam`), `Primary Type` = factor(`Primary Type`), `Primary Purpose` = factor(`Primary Purpose`), Region = factor(Region), `Modified Indicator` = factor(`Modified Indicator`), `Inspection Indicator` = factor(`Inspection Indicator`), `Assessment Indicator` = factor(`Assessment Indicator`), Spillway = factor(Spillway)

)


# Define the variables used for the training and predictions

predictors <- c(

  "Height (m)", "Hazard", "Assessment", "Regulated Dam", "Inspection Indicator", "Assessment Indicator", "Primary Type", "Primary Purpose", "Region", "Age", "Modified Indicator", "Spillway", "Distance to Nearest City (km)", "Inspection Frequency", "Length (km)", "Volume (m3)", "Surface (km2)", "Drainage (km2)"

)
```

```
X <- model.matrix( ~ . -1, data = dams[, predictors])
# Set target variable
y <- dams$`Probability of Failure`

set.seed(113)
# Random sampling for training and testing purpose
# 80% of the data will be used for training and the rest are used for testing
train_idx <- sample(1:nrow(X), 0.8 * nrow(X))
X_train <- X[train_idx, ]
y_train <- y[train_idx]
X_test <- X[-train_idx, ]
y_test <- y[-train_idx]

# Convert data to DMatrix format (optimized for XGBoost)
dtrain <- xgb.DMatrix(data = X_train, label = y_train)

# Define parameters
# Final parameters are determined using code (See Appendix: Hyperparameter Tuning) to ensure that the model will have the smallest error/ better performance
params <- list(
  objective = "reg:squarederror",  # For regression
  eta = 0.02,                      # Learning rate
  max_depth = 8,                   # Tree depth
  subsample = 0.8,                 # Fraction of data used per tree
  colsample_bytree = 0.8,          # Fraction of features used per tree
  gamma = 0,                       # Minimum loss reduction
  min_child_weight = 3             # Minimum sum of instance weight (hessian) in a child
)
```

```
# Train the model with the parameters set above
xgb_model <- xgb.train(
  params = params,
  data = dtrain,
  nrounds = 5000,                          # Number of rounds
  early_stopping_rounds = 30,      # Stop if no improvement after 30 rounds
  watchlist = list(train = dtrain)
)

# Predict on test data
pred <- predict(xgb_model, X_test)

# Calculate R-squared
rsq <- 1 - sum((y_test - pred)^2) / sum((y_test - mean(y_test))^2)
cat("R-squared:", rsq, "\n")

# If we want to visualize the effect of each factor in the prediction, we can run the code below:
library(xgboost)
importance_matrix <- xgb.importance(colnames(X), model = xgb_model)
xgb.plot.importance(importance_matrix)
```

Model 2: xgb_model_loss, used for total loss prediction

```
# Define the variables used for training and predictions

predictors_loss <- c("Height (m)", "Hazard", "Assessment", "Regulated Dam", "Inspection Indicator", "Assessment Indicator", "Primary Type", "Primary Purpose", "Region", "Age", "Modified Indicator", "Spillway", "Distance to Nearest City (km)", "Inspection Frequency", "Length (km)", "Volume (m3)", "Surface (km2)", "Drainage (km2)"

)


X_loss <- model.matrix( ~ . -1, data = dams[, predictors_loss])

# Define the target variable

y_loss <- dams$`Total Loss`


set.seed(128)

# Randomly sample 80% of the dams for model training, the rest are used to test the model

train_idx_loss <- sample(1:nrow(X_loss), 0.8 * nrow(X_loss))

X_train_loss <- X_loss[train_idx_loss, ]

y_train_loss <- y_loss[train_idx_loss]

X_test_loss <- X_loss[-train_idx_loss, ]

y_test_loss <- y_loss[-train_idx_loss]


# Convert data to DMatrix format (optimized for XGBoost)

dtrain_loss <- xgb.DMatrix(data = X_train_loss, label = y_train_loss)


# Define parameters

params_loss <- list(

  objective = "reg:squarederror",  # For regression

  eta = 0.01,                              # Learning rate

  max_depth = 10,                      # Tree depth

  subsample = 0.8,                    # Fraction of data used per tree

  colsample_bytree = 0.8          # Fraction of features used per tree

)
```

```r
# Train the model with the parameters defined
xgb_model_loss <- xgb.train(
  params_loss,
  data = dtrain_loss,
  nrounds = 10000,                    # Number of rounds
  early_stopping_rounds = 30,     # Stop if no improvement after 30 rounds
  watchlist = list(train = dtrain_loss)
)

# Predict on test data
pred_loss <- predict(xgb_model_loss, X_test_loss)

# Calculate R-squared
rsq_loss <- 1 - sum((y_test_loss - pred_loss)^2) / sum((y_test_loss - mean(y_test_loss))^2)
cat("R-squared:", rsq_loss, "\n")
```

Appendix: Hyperparameter Tuning

The following code to run is to determine the value for parameters that can yield the smallest Root Mean Squared Error (RMSE).

```
library(xgboost)

library(Matrix)

library(caret)


# Define the parameter grid for XGBoost tuning

# Can add other parameters here if needed

tune_grid <- expand.grid(

  nrounds = 500,

  eta = c(0.005, 0.01, 0.02),

  max_depth = c(8, 10, 12),

  gamma = 0,

  colsample_bytree = c(0.8, 1),

  min_child_weight = c(1, 3),

  subsample = c(0.8)

)


# Set up training control with 5-fold cross-validation

train_control <- trainControl(

  method = "cv",

  number = 5,

  verboseIter = TRUE

)
```

```
# Tune the model using caret's train function
set.seed(123)
xgb_caret_model <- train(
  x = X_train,
  y = y_train,
  trControl = train_control,
  tuneGrid = tune_grid,
  method = "xgbTree"
)

# Print the tuned model details
print(xgb_caret_model)

# Use the tuned model to predict on the test set
pred_tuned_prob <- predict(xgb_caret_model, newdata = X_test)

# Calculate R-squared for evaluation
rsq <- 1 - sum((y_test - pred_tuned_prob)^2) / sum((y_test - mean(y_test))^2)
cat("R-squared from caret-tuned model:", rsq, "\n")

# After we run the code above, cross-validation is performed for each combination of parameters in tuning grid, and then we selects the best model based on the chosen performance metric (by default for regression it's RMSE).
```

Appendix: Cox Ingersoll Rox Model

Inflation Rate Simulation

Parameter	Description	Value
κ (kappa)	The speed of reversion toward the long-term mean rate	0.189135
θ (theta)	The long-term mean rate	0.043005
σ (sigma)	Volatility	0.085537

```
library(readr)
library(tidyverse)
rates <- readr::read_csv("C:/Users/userc/Desktop/Interest Rates.csv")


# We set dt to be 1 since we are simulating annual rate
dt <- 1


negLogLikCIR_inflation <- function(params_inflation, r_inflation, dt) {
  # Extract parameters from the vector 'params_inflation'
  kappa_inflation <- params_inflation[1]  # Speed at which the rate reverts to the mean
  theta_inflation <- params_inflation[2]  # Long-term mean rate
  sigma_inflation <- params_inflation[3]  # Volatility of the rate

  # Ensure that all parameters are positive
  if (any(params_inflation <= 0)) return(1e10)
  # If not, return a high penalty (Penalty Method) so that the optimizer avoids negative parameters

  # Initialize the sum of the log-likelihood values
  logLikSum_inflation <- 0
```

```
  # Loop through the historical rates from the first to the second last rate available
  # Note: We now use r_inflation instead of r.
  for (i in 1:(length(r_inflation) - 1)) {
  # Scaling factor
    c_inflation <- (sigma_inflation^2 * (1 - exp(-kappa_inflation * dt))) / (4 * kappa_inflation)
  # Degrees of freedom
    df_inflation <- 4 * kappa_inflation * theta_inflation / sigma_inflation^2

  # Non-centrality parameter
    nc_inflation <- (4 * kappa_inflation * exp(-kappa_inflation * dt) * r_inflation[i]) / (sigma_inflation^2 *
(1 - exp(-kappa_inflation * dt)))

  # Ensure valid parameters for the chi-squared distribution (non negative)
    if (c_inflation <= 0 || df_inflation <= 0 || nc_inflation < 0) return(1e10)

  # Compute the density of the next rate under the scaled non-central chi-squared
    scaled_r_inflation <- r_inflation[i+1] / c_inflation

  # The PDF after taking log() on both side
    logLik_inflation <- dchisq(scaled_r_inflation, df = df_inflation, ncp = nc_inflation, log = TRUE) -
log(c_inflation)

    logLikSum_inflation <- logLikSum_inflation + logLik_inflation
  }

  # Return the negative log-likelihood for minimization.
  return(-logLikSum_inflation)
}

# Initial guesses for the parameters (kappa, theta, sigma)
init_params_inflation <- c(kappa_inflation = 0.5, theta_inflation = 0.05, sigma_inflation = 0.1)
```

```
# Optimize to find the best-fit parameters using historical inflation data
fit_inflation <- optim(par = init_params_inflation,
                  fn = negLogLikCIR_inflation,
                  r_inflation = rates$Inflation,   # Changed argument name
                  dt = dt,                          # The time step
                  method = "L-BFGS-B",        # A method that allows bounds
                  lower = c(0, 0, 0))               # Ensure parameters remain non-negative
# Extract the calibrated parameters
MLE_inflation <- fit_inflation$par
print(MLE_inflation)

# Use the last rate as the starting point
r0_inflation <- tail(rates$Inflation, 1)
nSteps <- 10
dt <- 1
MLE_kappa_inflation <- MLE_inflation[1]
MLE_theta_inflation <- MLE_inflation[2]
MLE_sigma_inflation <- MLE_inflation[3]

set.seed(223)
#CIR Simulation using the simulateCIR() function
simInflation <- simulateCIR(nSteps = nSteps, dt = dt, r0 = r0_inflation,
                     kappa = MLE_kappa_inflation,
                     theta = MLE_theta_inflation,
                     sigma = MLE_sigma_inflation)
plot(simInflation, type = "l",
    main = "Simulated CIR Inflation Rate",
    ylab = "Inflation Rate",
    xlab = "Time Step")
print(simInflation)
```

1-Year Risk-Free Rate Simulation

Parameter	Description	Value
κ (kappa)	The speed of reversion toward the long-term mean rate	0.072559
θ (theta)	The long-term mean rate	0.060408
σ (sigma)	Volatility	0.089110

```
# The process is almost identical
dt <- 1


negLogLikCIR_RiskFree <- function(params_RiskFree, r_RiskFree, dt) {
  kappa_RiskFree <- params_RiskFree[1]  # Speed at which the rate reverts to the mean
  theta_RiskFree <- params_RiskFree[2]  # Long-term mean rate
  sigma_RiskFree <- params_RiskFree[3]  # Volatility of the rate


 if (any(params_RiskFree <= 0)) return(1e10)
 logLikSum_RiskFree <- 0


  for (i in 1:(length(r_RiskFree) - 1)) {
    # Scaling factor
    c_RiskFree <- (sigma_RiskFree^2 * (1 - exp(-kappa_RiskFree * dt))) / (4 * kappa_RiskFree)
    # Degrees of freedom
    df_RiskFree <- 4 * kappa_RiskFree * theta_RiskFree / sigma_RiskFree^2
    # Non-centrality parameter
    nc_RiskFree <- (4 * kappa_RiskFree * exp(-kappa_RiskFree * dt) * r_RiskFree[i]) /
(sigma_RiskFree^2 * (1 - exp(-kappa_RiskFree * dt)))


    if (c_RiskFree <= 0 || df_RiskFree <= 0 || nc_RiskFree < 0) return(1e10)


    scaled_r_RiskFree <- r_RiskFree[i+1] / c_RiskFree
    logLik_RiskFree <- dchisq(scaled_r_RiskFree, df = df_RiskFree, ncp = nc_RiskFree, log = TRUE) -
log(c_RiskFree)
```

```
    logLikSum_RiskFree <- logLikSum_RiskFree + logLik_RiskFree
  }

  return(-logLikSum_RiskFree)
}

init_params_RiskFree <- c(kappa_RiskFree = 0.5, theta_RiskFree = 0.05, sigma_RiskFree = 0.1)

fit_RiskFree <- optim(par = init_params_RiskFree,
                      fn = negLogLikCIR_RiskFree,
                      r_RiskFree = rates$`1-yr Risk Free Annual Spot Rate`,
                      dt = dt,
                      method = "L-BFGS-B",
                      lower = c(0, 0, 0))

MLE_RiskFree <- fit_RiskFree$par
print(MLE_RiskFree)
r0_RiskFree <- tail(rates$`1-yr Risk Free Annual Spot Rate`, 1)

# Simulation settings
nSteps <- 10
dt <- 1

MLE_kappa_RiskFree <- MLE_RiskFree[1]
MLE_theta_RiskFree <- MLE_RiskFree[2]
MLE_sigma_RiskFree <- MLE_RiskFree[3]
```

```
set.seed(223)

simRiskFree <- simulateCIR(nSteps = nSteps, dt = dt, r0 = r0_RiskFree,
                           kappa = MLE_kappa_RiskFree,
                           theta = MLE_theta_RiskFree,
                           sigma = MLE_sigma_RiskFree)

plot(simRiskFree, type = "l",
     main = "Simulated CIR 1-yr_Risk_Free_Annual_Spot_Rate",
     ylab = "1-yr_Risk_Free_Annual_Spot_Rate",
     xlab = "Time Step")

print(simRiskFree)

# Writing the simulated rates into a new CSV file
# Number of simulated rates
n <- length(simRiskFree)

# Create a sequence of years starting from 2024
years <- 2024 + 0:(n-1)

# Create a data frame with two columns: Year, Inflation and Spot Rate
simulated_rates <- data.frame(
  Year = years,
  Inflation = simInflation,
  `1-yr_Risk_Free_Annual_Spot_Rate` = simRiskFree
)

write.csv(simulated_rates, "C:/Users/userc/Desktop/Simulated_Rates.csv", row.names = FALSE)
```

Appendix: Monte Carlo Simulation for Coverage Limit

```
library(dplyr)

library(caret)

library(readr)


dams <- readr::read_csv("C:/Users/userc/Desktop/Insured List.csv")

results_list <- list()

n_simulations <- 50000


# Create a one-dam simulation by sampling one value per variable.

# Use check.names = FALSE to preserve the column names exactly.

for (i in 1:n_simulations) {

dam_sim <- data.frame(

    "Height (m)" = sample(dams$`Height (m)`, 1, replace = TRUE),

    "Volume (m3)" = sample(dams$`Volume (m3)`, 1, replace = TRUE),

    "Length (km)" = sample(dams$`Length (km)`, 1, replace = TRUE),

    "Surface (km2)" = sample(dams$`Surface (km2)`, 1, replace = TRUE),

    "Drainage (km2)" = sample(dams$`Drainage (km2)`, 1, replace = TRUE),

    "Inspection Frequency" = sample(dams$`Inspection Frequency`, 1, replace = TRUE),

    "Distance to Nearest City (km)" = sample(dams$`Distance to Nearest City (km)`, 1, replace = TRUE),

    "Age" = sample(dams$Age, 1, replace = TRUE),

    check.names = FALSE

  )


  # Add categorical variables with full factor level specifications (as the trained model to avoid error
message)

  dam_sim$Region <- factor(

    sample(dams$Region, size = 1, replace = TRUE),

    levels = c("Flumevale", "Lyndrassia", "Navaldia")

  )
```

```
dam_sim$`Regulated Dam` <- factor(
  sample(dams$`Regulated Dam`, size = 1, replace = TRUE),
  levels = c("Yes", "No")
)

dam_sim$`Primary Purpose` <- factor(
  sample(dams$`Primary Purpose`, size = 1, replace = TRUE),
  levels = c("Debris Control", "Irrigation", "Recreation",
             "Flood Risk Reduction", "Fire Protection, Stock, Or Small Fish Pond",
             "Fish and Wildlife Pond", "Water Supply", "Hydroelectric", "Other",
             "Tailings", "Grade Stabilization", "Navigation")
)

dam_sim$Hazard <- factor(
  sample(dams$Hazard, size = 1, replace = TRUE),
  levels = c("Low", "High", "Significant", "Undetermined")
)

dam_sim$Assessment <- factor(
  sample(dams$Assessment, size = 1, replace = TRUE),
  levels = c("Satisfactory", "Fair", "Poor", "Unsatisfactory", "Not Rated", "Not Available")
)

dam_sim$`Primary Type` <- factor(
  sample(dams$`Primary Type`, size = 1, replace = TRUE),
  levels = c("Arch", "Buttress", "Concrete", "Earth", "Gravity",
             "Masonry", "Multi-Arch", "Other", "Rockfill",
             "Roller-Compacted Concrete", "Stone", "Timber Crib")
)
```

```
dam_sim$`Modified Indicator` <- factor(
  sample(dams$`Modified Indicator`, size = 1, replace = TRUE),
  levels = c("Yes", "No")
)

dam_sim$`Assessment Indicator` <- factor(
  sample(dams$`Assessment Indicator`, size = 1, replace = TRUE),
  levels = c("Yes", "No")
)

dam_sim$`Inspection Indicator` <- factor(
  sample(dams$`Inspection Indicator`, size = 1, replace = TRUE),
  levels = c("Yes", "No")
)

dam_sim$Spillway <- factor(
  sample(dams$Spillway, size = 1, replace = TRUE),
  levels = c("Controlled", "Uncontrolled")
)

# Debug: Check if all predictors exist in dam_sim
missing_predictors <- setdiff(predictors_loss, names(dam_sim))
if (length(missing_predictors) > 0) {
  stop("Iteration ", i, ": The following predictors are missing in dam_sim: ",
       paste(missing_predictors, collapse = ", "))
}
```

```
# Manually build the model matrix
  X_sim_loss <- model.matrix( ~ . - 1, data = dam_sim[, predictors_loss],
 contrasts.arg = list(
 Region = contr.treatment(c("Flumevale", "Lyndrassia", "Navaldia"), base = 1),
 `Regulated Dam` = contr.treatment(c("Yes", "No"), base = 1),
 `Primary Purpose` = contr.treatment(c("Debris Control", "Irrigation", "Recreation",
                               "Flood Risk Reduction", "Fire Protection, Stock, Or Small Fish Pond",
                                "Fish and Wildlife Pond", "Water Supply", "Hydroelectric", "Other",
                                "Tailings", "Grade Stabilization", "Navigation"), base = 1),
  Hazard = contr.treatment(c("Low", "High", "Significant", "Undetermined"), base = 1),
   Assessment = contr.treatment(c("Satisfactory", "Fair", "Poor", "Unsatisfactory", "Not Rated",   "Not
Available"), base = 1),
  `Primary Type` = contr.treatment(c("Arch", "Buttress", "Concrete", "Earth", "Gravity",
                                 Masonry", "Multi-Arch", "Other", "Rockfill",
                                "Roller-Compacted Concrete", "Stone", "Timber Crib"), base = 1))
  )
```

We will need to manually build the model matrix if the model is applied on a dataset that does not have some of the variables from the original dataset used to train the model.

For example, a dataset:

Dam	Region	Hazard
1	A	Low
2	B	High
3	A	Low

The model.matrix will look like:

Dam	Region A	Region B	Low	High
1	1	0	1	0
2	0	1	0	1
3	1	0	1	0

If we apply it to a new dataset:

Dam	Region	Hazard
1	A	Low
2	B	Low
3	A	Low

The model.matrix will face errors since:

Dam	Region A	Region B	Low	High
1	1	0	1	0
2	0	1	1	0
3	1	0	1	0

The error message will pop up, we have to make sure that the model matrix has the same structure as during training.

```
# Ensure that simulation data matrix (X_sim_loss) has exactly the same columns as the training data
matrix (X_train_loss)

 missing_cols <- setdiff(colnames(X_train_loss), colnames(X_sim_loss))

 if (length(missing_cols) > 0) {

   X_sim_loss <- cbind(X_sim_loss,

                       matrix(0, nrow = nrow(X_sim_loss), ncol = length(missing_cols),

                              dimnames = list(NULL, missing_cols)))

 }

 X_sim_loss <- X_sim_loss[, colnames(X_train_loss), drop = FALSE]
```

```r
# Apply the predictive model
  predicted_loss <- predict(xgb_model_loss, X_sim_loss)
  dam_sim$PredictedLoss <- predicted_loss

  results_list[[i]] <- dam_sim
}

# Combine all simulated dam data into one data frame for the next step
final_results <- do.call(rbind, results_list)

# Post-Simulation Analysis (Grouped according to Regions and Purpose)
# Each Purpose in Different Region have their own coverage limit (70th percentile)
output_dir <- "C:/Users/userc/Desktop/Monte Carlo"

summary_table <- final_results %>%
  filter(
    `Regulated Dam` == "Yes", !Assessment %in% c("Not Rated", "Not Available"), Hazard != "Undetermined", `Primary Type` == "Earth"
  ) %>%
  group_by(Region, `Primary Purpose`) %>%
  summarize( CoverageLimit = quantile(PredictedLoss, 0.70, na.rm = TRUE), .groups = 'drop'
  )
write_csv(summary_table, file.path(output_dir, "Coverage_Limits_Summary.csv"))

combinations <- final_results %>%
  distinct(Region, `Primary Purpose`) %>%
  drop_na()
```

```
# For each unique combinations of regions and purpose, apply filter and store to different csv files
for (i in 1:nrow(combinations)) {
  current_region <- combinations$Region[i]
  current_purpose <- combinations$`Primary Purpose`[i]

  subset_df <- final_results %>%
filter( Region == current_region, `Primary Purpose` == current_purpose, `Regulated Dam` ==
"Yes", !Assessment %in% c("Not Rated", "Not Available"), Hazard != "Undetermined", `Primary Type`
== "Earth"
    )

  if (nrow(subset_df) > 0) {
    safe_purpose <- gsub("[^A-Za-z0-9]", "_", current_purpose)
    csv_filename <- file.path(
      output_dir,
      paste0("Simulated_Dams_", current_region, "_", safe_purpose, ".csv")
    )
    write_csv(subset_df, csv_filename)
    cat("Saved:", csv_filename, "\n")
  }
}

print(summary_table, n=50)
```

Appendix: Underwriting

```
library(dplyr)


dams <- readr::read_csv("C:/Users/userc/Desktop/DamData_Rated.csv")


# Criteria 1

criteria1_dams <- dams %>%

  filter( `Regulated Dam` == "Yes", `Primary Type` == "Earth", `Hazard` %in% c("Low", "High", "Significant"), `Assessment` %in% c("Satisfactory", "Fair", "Poor", "Unsatisfactory"), `Inspection Indicator` == "No", `Assessment Indicator` == "No")

count_criteria1 <- nrow(criteria1_dams)

print(count_criteria1)


# Criteria 2

criteria2_dams <- dams %>%

  filter( `Regulated Dam` == "Yes", `Primary Type` == "Earth", `Hazard` %in% c("Low", "High", "Significant"), `Assessment` %in% c("Satisfactory", "Fair", "Poor", "Unsatisfactory"),

    # OR Logical Function

    (`Inspection Indicator` == "Yes" | `Assessment Indicator` == "Yes")

  )

count_watchlist <- nrow(criteria2_dams)

print(count_watchlist)


# Pick the dams below the 90th percentile

quantile_threshold2 <- quantile(criteria2_dams$final_rating, 0.9, na.rm = TRUE)

criteria2_insured_dams <- criteria2_dams %>%

 filter(final_rating <= quantile_threshold2)

print(nrow(criteria2_insured_dams))
```

```
# Criteria 3
criteria3_dams_1 <- dams %>%
  filter( `Primary Type` == "Earth", `Regulated Dam` == "Yes", `Hazard` %in% c("Undetermined"),
`Assessment` %in% c("Satisfactory", "Fair", "Poor", "Unsatisfactory")
  )
print(nrow(criteria3_dams_1))

criteria3_dams_2 <- dams %>%
  filter( `Primary Type` == "Earth", `Regulated Dam` == "Yes", `Hazard` %in% c("Low", "High",
"Significant"), `Assessment` %in% c("Not Rated","Not Available")
  )
print(nrow(criteria3_dams_2))

# Select the dams below 70th percentile
quantile_threshold3_1 <- quantile(criteria3_dams_1$final_rating, 0.7, na.rm = TRUE)
criteria3_insured_dams_1 <- criteria3_dams_1 %>%
  filter(final_rating <= quantile_threshold3_1)

quantile_threshold3_2 <- quantile(criteria3_dams_2$final_rating, 0.7, na.rm = TRUE)
criteria3_insured_dams_2 <- criteria3_dams_2 %>%
  filter(final_rating <= quantile_threshold3_2)

print(nrow(criteria3_insured_dams_1))
print(nrow(criteria3_insured_dams_2))

hazard_values <- c("High", "Significant", "Low")
assessment_values <- c("Satisfactory", "Fair", "Poor", "Unsatisfactory")
```

```
# Randomly assign a hazard and assessment classification for the dams that have hazard classification of "Undetermined" and assessment classification of "Not Rated" and "Not Available" after they have passed the required inspection for financial projection purpose.

criteria3_insured_dams_1$Hazard <- sample(hazard_values, size = nrow(criteria3_insured_dams_1), replace = TRUE)

criteria3_insured_dams_1$Assessment <- sample(assessment_values, size = nrow(criteria3_insured_dams_1), replace = TRUE)


criteria3_insured_dams_2$Hazard <- sample(hazard_values, size = nrow(criteria3_insured_dams_2), replace = TRUE)

criteria3_insured_dams_2$Assessment <- sample(assessment_values, size = nrow(criteria3_insured_dams_2), replace = TRUE)


# Criteria 4

regulated_to_non <- dams %>%

  filter( `Primary Type` == "Earth", `Regulated Dam` == "Yes", Hazard == "Undetermined", Assessment %in% c("Not Rated", "Not Available")

  )

print(nrow(regulated_to_non))


# Non Criteria 1

non_regulated_dams <- dams %>%

  filter(`Regulated Dam` == "No", `Primary Type` == "Earth",)


non_regulated_list <- bind_rows(non_regulated_dams, regulated_to_non)


# We randomly pick 50% of the non-regulated dams assuming they have pass the required inspection for financial projection purpose.

set.seed(123)

sample_size <- floor(0.5 * nrow(non_regulated_list))

sample_indices <- sample(seq_len(nrow(non_regulated_list)), size = sample_size)
```

```
non_regulated_passed <- non_regulated_list[sample_indices, ]
non_regulated_passed[["Regulated Dam"]] <- "Yes"

print(nrow(non_regulated_passed))

non_regulated_passed$`Regulated Dam` <- "Yes"
non_criteria1_dams <- non_regulated_passed %>%
  filter( `Hazard` %in% c("Low", "High", "Significant"), `Assessment` %in% c("Satisfactory", "Fair", "Poor", "Unsatisfactory"), `Inspection Indicator` == "No", `Assessment Indicator` == "No"
  )
print(nrow(non_criteria1_dams))
print(nrow(non_regulated_passed))

# Non Criteria 2
non_criteria2_dams <- non_regulated_passed %>%
  filter( `Hazard` %in% c("Low", "High", "Significant"), `Assessment` %in% c("Satisfactory", "Fair", "Poor", "Unsatisfactory"),
    # OR Logical Function
    (`Inspection Indicator` == "Yes" | `Assessment Indicator` == "Yes")
  )
count_watchlist <- nrow(non_criteria2_dams)
print(count_watchlist)

# We select the dams that have hazard index below 90th percentile
quantile_threshold_non2 <- quantile(non_criteria2_dams$final_rating, 0.9, na.rm = TRUE)
non_criteria2_insured_dams <- non_criteria2_dams %>%
  filter(final_rating <= quantile_threshold_non2)
print(nrow(non_criteria2_insured_dams))
```

```
# Non Criteria 3

non_criteria3_dams_1 <- non_regulated_passed %>%
  filter( `Primary Type` == "Earth", `Regulated Dam` == "Yes",  `Hazard` == "Undetermined",
`Assessment` %in% c("Satisfactory", "Fair", "Poor", "Unsatisfactory")
  )
print(nrow(non_criteria3_dams_1))


non_criteria3_dams_2 <- non_regulated_passed %>%
  filter( `Primary Type` == "Earth", `Regulated Dam` == "Yes", `Hazard` %in% c("Low", "High",
"Significant"), `Assessment` %in% c("Not Rated", "Not Available")
  )
print(nrow(non_criteria3_dams_2))


# We select the dams that have hazard index below 70th percentile
quantile_threshold_non_3_1 <- quantile(non_criteria3_dams_1$final_rating, 0.7, na.rm = TRUE)
non_criteria3_insured_dams_1 <- non_criteria3_dams_1 %>%
  filter(final_rating <= quantile_threshold_non_3_1)


quantile_threshold_non_3_2 <- quantile(non_criteria3_dams_2$final_rating, 0.7, na.rm = TRUE)
non_criteria3_insured_dams_2 <- non_criteria3_dams_2 %>%
  filter(final_rating <= quantile_threshold_non_3_2)
print(nrow(non_criteria3_insured_dams_1))
print(nrow(non_criteria3_insured_dams_2))


hazard_values <- c("High", "Significant", "Low")
assessment_values <- c("Satisfactory", "Fair", "Poor", "Unsatisfactory")


non_criteria3_insured_dams_1$Hazard <- sample(hazard_values, size =
nrow(non_criteria3_insured_dams_1), replace = TRUE)
non_criteria3_insured_dams_1$Assessment <- sample(assessment_values, size =
nrow(non_criteria3_insured_dams_1), replace = TRUE)
```

```
non_criteria3_insured_dams_2$Hazard <- sample(hazard_values, size =
nrow(non_criteria3_insured_dams_2), replace = TRUE)

non_criteria3_insured_dams_2$Assessment <- sample(assessment_values, size =
nrow(non_criteria3_insured_dams_2), replace = TRUE)


# Non Criteria 4

non_criteria4_dams <- non_regulated_passed %>%

  filter(Hazard == "Undetermined",  Assessment %in% c("Not Rated", "Not Available")

  )


quantile_threshold_non_4 <- quantile(non_criteria4_dams$final_rating, 0.5, na.rm = TRUE)

non_criteria4_insured_dams <- non_criteria4_dams %>%

  filter(final_rating <= quantile_threshold_non_4)


hazard_values <- c("High", "Significant", "Low")

assessment_values <- c("Satisfactory", "Fair", "Poor", "Unsatisfactory")


non_criteria4_insured_dams$Hazard <- sample(hazard_values, size = nrow(non_criteria4_insured_dams),
replace = TRUE)

non_criteria4_insured_dams$Assessment <- sample(assessment_values, size =
nrow(non_criteria4_insured_dams), replace = TRUE)


print(nrow(non_criteria4_dams))

print(nrow(non_criteria4_insured_dams))


# Combine all the individual list

criteria3_insured_dams <- bind_rows(criteria3_insured_dams_1, criteria3_insured_dams_2)


insured_dam_list_regulated <- bind_rows( criteria1_dams, criteria2_insured_dams,
criteria3_insured_dams

) %>%

  distinct()
```

```
cat("Number of insured regulated dams:", nrow(insured_dam_list_regulated), "\n")

non_criteria3_insured_dams <- bind_rows(non_criteria3_insured_dams_1,
non_criteria3_insured_dams_2)

insured_dam_list_non_regulated <- bind_rows( non_criteria1_dams, non_criteria2_insured_dams,
non_criteria3_insured_dams, non_criteria4_insured_dams
) %>%
  distinct()

cat("Number of insured non-regulated dams:", nrow(insured_dam_list_non_regulated), "\n")

final_insured_dam_list <- bind_rows(insured_dam_list_regulated,
insured_dam_list_non_regulated) %>% distinct()

cat("Total number of insured dams:", nrow(final_insured_dam_list), "\n")

library(readr)
write_csv(final_insured_dam_list, "C:/Users/userc/Desktop/Insured List.csv")
```

Appendix: The Computation of Annual Premium

```
library(dplyr)
library(readr)
library(fitdistrplus)

dams <- read_csv("C:/Users/userc/Desktop/DamLoss_Scenario.csv")
rates <- read_csv("C:/Users/userc/Desktop/Simulated_Rates.csv")

table(dams$Hazard)

# Compute the adjustment factor for cashflow at the beginning of  xth year
n_years <- 10
adjustment_factors <- numeric(n_years)
adjustment_factors[1] <- 1
for(t in 1:(n_years - 1)) {
  cumulative_inflation <- prod(1 + rates$Inflation[1:t])
  cumulative_interest  <- prod(1 + rates$`1.yr_Risk_Free_Annual_Spot_Rate`[1:t])
  adjustment_factors[t + 1] <- cumulative_inflation / cumulative_interest
}

# Nominal costs for inspection and service (in Qm)
dams <- dams %>%
  mutate(
    Inspection_Cost_nominal = case_when(
      Hazard == "Low" ~ 0.02576335,
      Hazard == "Significant" ~ 0.02962785,
      Hazard == "High" ~ 0.03349235,
      TRUE ~ NA_real_
    ),
```

```
    Service_Cost = case_when(
      Height_Class == 1 & Service_Needed == "Repair" ~ 0.400,
      Height_Class == 1 & Service_Needed == "Retrofit" ~ 1.380,
      Height_Class == 1 & Service_Needed == "Rehabilitation" ~ 2.870,
      Height_Class == 2 & Service_Needed == "Repair" ~ 0.700,
      Height_Class == 2 & Service_Needed == "Retrofit" ~ 1.800,
      Height_Class == 2 & Service_Needed == "Rehabilitation" ~ 2.570,
      Height_Class == 3 & Service_Needed == "Repair" ~ 1.100,
      Height_Class == 3 & Service_Needed == "Retrofit" ~ 1.803,
      Height_Class == 3 & Service_Needed == "Rehabilitation" ~ 2.547,
      Height_Class == 4 & Service_Needed == "Repair" ~ 1.400,
      Height_Class == 4 & Service_Needed == "Retrofit" ~ 2.580,
      Height_Class == 4 & Service_Needed == "Rehabilitation" ~ 3.620,
      Height_Class == 5 & Service_Needed == "Repair" ~ 2.000,
      Height_Class == 5 & Service_Needed == "Retrofit" ~ 2.938,
      Height_Class == 5 & Service_Needed == "Rehabilitation" ~ 4.780,
      Height_Class == 6 & Service_Needed == "Repair" ~ 2.600,
      Height_Class == 6 & Service_Needed == "Retrofit" ~ 5.294,
      Height_Class == 6 & Service_Needed == "Rehabilitation" ~ 9.500,
      TRUE ~ 0
    )
  )

# Coverage Limit
coverage_table <- tribble(
  ~Region,      ~PrimaryPurpose, ~CoverageLimit,
  "Flumevale",  "Debris Control", 639.93,
  "Flumevale",  "Irrigation", 606.16,
  "Flumevale",  "Recreation", 596.57,
  "Flumevale",  "Flood Risk Reduction", 586.49,
```

```
  "Flumevale",  "Fire Protection, Stock, Or Small Fish Pond", 598.50,
  "Flumevale",  "Fish and Wildlife Pond", 581.54,
  "Flumevale",  "Water Supply", 612.85,
  "Flumevale",  "Hydroelectric", 634.04,
  "Flumevale",  "Other", 545.55,
  "Flumevale",  "Tailings", 652.95,
  "Lyndrassia", "Debris Control", 578.31,
  "Lyndrassia", "Irrigation", 611.33,
  "Lyndrassia", "Recreation", 597.39,
  "Lyndrassia", "Flood Risk Reduction", 584.32,
  "Lyndrassia", "Fire Protection, Stock, Or Small Fish Pond", 590.94,
  "Lyndrassia", "Fish and Wildlife Pond", 553.75,
  "Lyndrassia", "Water Supply", 612.73,
  "Lyndrassia", "Hydroelectric", 579.04,
  "Lyndrassia", "Other", 576.58,
  "Lyndrassia", "Tailings", 581.53,
  "Lyndrassia", "Grade Stabilization", 564.96,
  "Navaldia",   "Debris Control", 440.69,
  "Navaldia",   "Irrigation", 516.27,
  "Navaldia",   "Recreation", 539.85,
  "Navaldia",   "Flood Risk Reduction", 506.62,
  "Navaldia",   "Fire Protection, Stock, Or Small Fish Pond", 513.92,
  "Navaldia",   "Fish and Wildlife Pond", 576.82,
  "Navaldia",   "Water Supply", 526.06,
  "Navaldia",   "Hydroelectric", 570.55,
  "Navaldia",   "Other", 512.30,
  "Navaldia",   "Tailings", 598.61,
  "Navaldia",   "Grade Stabilization", 533.39
)
```

Add columns like coverage_rate, expected_payout, annual_prob_with (the predicted probability) to make the calculations easier and for error checking purpose

```
dams_merged <- dams %>%
  left_join(coverage_table, by = c("Region" = "Region", "Primary Purpose" = "PrimaryPurpose")) %>%
  mutate(
    coverage_rate = case_when(
      rating_groups == "Low" ~ 1.0,
      rating_groups == "Medium" ~ 0.8,
      rating_groups == "High" ~ 0.6,
      TRUE ~ NA_real_
    ),
    expected_payout = pmin(predicted_total_loss * coverage_rate, CoverageLimit, na.rm = TRUE),
    annual_prob_with = 1 - (1 - predicted_probability)^(1/10),
    risk_factor = case_when(
      rating_groups == "High" ~ 1.2,
      rating_groups == "Medium" ~ 1.1,
      rating_groups == "Low" ~ 1.0,
      TRUE ~ 1.0
    )
  )
```

Compute the PV of expected loss at the beginning of the each policy year and then store them into a matrix, the total PV of each individual dam throughout the policy period will be the row sum of that matrix.

```
expected_loss_without <- matrix(0, nrow = nrow(dams_merged), ncol = n_years)
expected_loss_with <- matrix(0, nrow = nrow(dams_merged), ncol = n_years)
for(t in 1:n_years) {
  cumulative_inflation <- prod(rates$Inflation[1:t] + 1)
  cumulative_interest  <- prod(rates$`1.yr_Risk_Free_Annual_Spot_Rate`[1:t] + 1)
  adjustment_factor <- cumulative_inflation / cumulative_interest
```

```
  expected_loss_with[, t] <- dams_merged$expected_payout * dams_merged$annual_prob_with * 
dams_merged$risk_factor * adjustment_factor
}

dams_merged$exp_loss_with <- rowSums(expected_loss_with)

# The value of contingent annuity due is actually the sum of adjustment factors from each policy year
annuity_due_with <- sapply(dams_merged$annual_prob_with, function(p) {
  sum(adjustment_factors * (1 - p)^(0:(n_years - 1)))
})

# Calculation APV for each dam using the function we defined
APV_inspection <- function(base_cost, freq, ann_prob, n_years, adj_factors) {
  n_terms <- floor(n_years / freq)
  sum(sapply(0:(n_terms - 1), function(t) {
    base_cost * (1 - ann_prob)^(freq * t) * adj_factors[1 + freq * t]
  }))
}

# Different hazards have different amount, so we need to adjust the function input accordingly
dams_merged <- dams_merged %>%
  rowwise() %>%
  mutate(
    APV_Inspection = case_when(
      Hazard == "High" ~ APV_inspection(0.03349235, 1, annual_prob_with, n_years, 
adjustment_factors),
      Hazard == "Significant" ~ APV_inspection(0.02962785, 2, annual_prob_with, n_years, 
adjustment_factors),
      Hazard == "Low" ~ APV_inspection(0.02576335, 5, annual_prob_with, n_years, adjustment_factors),
      TRUE ~ NA_real_
    )
```

```
  ) %>%
  ungroup()

dams_merged <- dams_merged %>%
  mutate(
    APV_Service = case_when(
      Service_Needed == "Rehabilitation" ~ Service_Cost,
      Service_Needed == "Retrofit" ~ Service_Cost * adjustment_factors[3] * (1 - annual_prob_with)^2,
      Service_Needed == "Repair" ~ Service_Cost * adjustment_factors[5] * (1 - annual_prob_with)^4,
      TRUE ~ 0
    )
  )

dams_merged <- dams_merged %>%
  mutate(total_exp_loss = exp_loss_with + APV_Inspection + APV_Service)

# The reinsurance loading factor
# The concept here is the Equipvalence Principle
dams_merged <- dams_merged %>%
  mutate(annual_rate = total_exp_loss*1.05/annuity_due_with)

write_csv(dams_merged, "C:/Users/userc/Desktop/Insured List With Loss.csv")
```

Appendix: Monte Carlo Simulation for Dam Failure

```
library(dplyr)

library(readr)

library(fitdistrplus)


base_data <- read_csv("C:/Users/userc/Desktop/Insured List With Loss.csv")

output_dir <- "C:/Users/userc/Desktop/Projection"


for (yr in 1:10) {

  current_data <- if (yr == 1) base_data else good_data

  # Obtain the shape parameter values after fitting the annual probability of failure into the beta
distribution

  fit_beta <- fitdist(current_data$annual_prob_with, "beta")

  set.seed(123 + yr)


  current_data <- current_data %>%

    rowwise() %>%

    mutate(

      sim_prob = rbeta(1, shape1 = fit_beta$estimate["shape1"], shape2 = fit_beta$estimate["shape2"]),

      fail = if_else(rbinom(1, size = 1, prob = sim_prob) == 1, "Yes", "No")

    ) %>%

    ungroup()


  failed_data <- filter(current_data, fail == "Yes")

  good_data   <- filter(current_data, fail == "No")

  write_csv(failed_data, paste0(output_dir, "/failed", yr, ".csv"))

  write_csv(good_data, paste0(output_dir, "/good", yr, ".csv"))

}
```

References

Ageing dams pose growing threat: UN. (2021, January 26). PreventionWeb. https://www.preventionweb.net/news/ageing-dams-pose-growing-threat-un

APMonitor Optimization Suite. (n.d.). *XGBoost regressor*. https://apmonitor.com/pds/index.php/Main/XGBoostRegressor

ASDSO Dam Safety Toolbox. (n.d.). *Flood risk management.* https://damtoolbox.org/wiki/Flood_Risk_Management

ASDSO Dam Safety Toolbox. (n.d.). *Recreation.* https://damtoolbox.org/wiki/Recreation

ASDSO Dam Safety Toolbox. (n.d.). *Water Supply & Regulation.* https://damtoolbox.org/wiki/Water_Supply_%26_Regulation

Association of State Dam Safety Officials. (2025). *The cost of rehabilitating dams in the U.S.* https://damsafety-prod.s3.amazonaws.com/s3fs-public/files/2025%20ASDSO%20Costs%20of%20Dam%20Rehab%20Report.pdf

Carrieann. (2024, November 21). *Engineering booms as demand for dams surges.* International Water Power and Dam Construction. https://www.waterpowermagazine.com/analysis/engineering-booms-as-demand-for-dams-surges/?cf-view

Central Water Commission. (2018). *Manual for rehabilitation of large dams.* (CDSO_MAN_DS_02_v1.0). https://damsafety.cwc.gov.in/ecm-includes/PDFs/Manual_for_Rehabilitation_of_Large_Dams.pdf

Department of Irrigation and Drainage Malaysia. (2017). *Malaysia dam safety management (MyDAMS)*. https://www.water.gov.my/jps/resources/PDF/MyDAMS_2017_(Free_Copy).pdf

Federal Emergency Management Agency. (2013). *Federal guidelines for inundation mapping of flood risks associated with dam incidents and failures.* (FEMA P-946). https://www.fema.gov/sites/default/files/2020-08/fema_dam-safety_inundation-mapping-flood-risks.pdf

Hoellein, T. J., Kim, L. H., Lazcano, R. F., & Vincent, A. E. S. (2024). Debris dams retain trash, mostly plastic, in urban streams. *Freshwater Science*, *43*(1), 94–106. https://doi.org/10.1086/729305

Hughson, L. A. (2023, December 1). *How dams reduce the impact of floods and fires.* Big Ditch Dam Building Company. https://www.bigditch.com.au/how-dams-reduce-the-impact-of-floods-and-fires/?srsltid=AfmBOoo79MhYtARwooNnk5-PhiDsvnS_iPJKReQpCytaRZyhYjt6wO63

IndustriALL. (2019, February 13). *What are tailings dams?* https://www.industriall-union.org/why-you-need-to-know-about-tailings-dams

Kenton, W. (2023, September 25). *Cox-Ingersoll-Ross Model (CIR): Overview, formula, and limitations*. Investopedia. https://www.investopedia.com/terms/c/cox-ingersoll-ross-model.asp

Montana Department of Natural Resources & Conservation Dam Safety Program. (2025, March 5). *A pilot study of dam inspection expense assistance.* https://dnrc.mt.gov/_docs/water/Dam_Safety/PUBLICATIONS/Bulletins/DNRC-Report-Pilot-Study-of-Dam-Inspection-Expense-Assistance_2025-03-05.pdf

Norwegian Refuge Council. (2024, September 10). *One year on: Looking back at Derna's flood disaster.* https://www.nrc.no/perspectives/2024/one-year-on-looking-back-at-dernas-flood-disaster

RDocumentation. (n.d.). *mice: Multivariate imputation by chained equations.* https://www.rdocumentation.org/packages/mice/versions/3.17.0/topics/mice

Red River Retention Authority. *Parts of a dam*. (n.d.). https://www.redriverretentionauthority.net/parts-of-a-dam.html

The Constructor. (2018, August 29). *Types of earthfill dams – Applications and advantages*. https://theconstructor.org/water-resources/earthfill-dams-its-classification/2273/

United States Department of Agriculture. (2020). *Grade stabilization structure.* (410-CPS). https://www.nrcs.usda.gov/sites/default/files/2022-09/Grade_Stabilization_Structure_410_CPS_10_2020.pdf

United States Department of Agriculture. (2020). *Irrigation reservoir.* (436-CPS). https://www.nrcs.usda.gov/sites/default/files/2022-09/Irrigation_Reservoir_436_CPS_9_2020.pdf

University of Virginia Library. (n.d.). *Getting started with multiple imputation in R.* https://library.virginia.edu/data/articles/getting-started-with-multiple-imputation-in-r

US Army Corps of Engineers. (2004). *General design and construction considerations for earth and rock-fill dams.* (EM 1110-2-2300). https://www.publications.usace.army.mil/portals/76/publications/engineermanuals/em_1110-2-2300.pdf

Vollebergh, H., & Dijk, J. (2016). *Taxes and fees of regional water authorities in the Netherlands*. Institute European Environmental Policy. https://ieep.eu/wp-content/uploads/2022/12/NL-Water-Taxes-Fees-final.pdf