Incorporating climate change in I&M planning for engineering structures

A Deep Reinforcement Learning Framework

by

Sammie Knoppert

Student Name	Sammie Knoppert
Student Number	4972961
Date	4 May 2023

Student number: 4972961
Project duration: May 5, 2022 – May 5, 2023
Thesis committee: Dr. C. Andriotis, TU Delft, Daily supervisor
Dr. ir. M. Ottele, TU Delft, Chair
Dr. S. L. M. Lhermitte, TU Delft

TU Delft

Preface

Writing the last lines of this thesis brings a mix of feelings to overcome me. First, a great sense of relief for the struggles experienced to achieve this milestone coming to an end. Second, I have a hint of sadness since this chapter of my life which I enjoyed immensely and have had the opportunity to establish deep friendships and meet great people comes to an end. And finally, I feel very motivated to start working in this exciting field of Civil Engineering and start my 'working life.'

I want to sincerely thank my supervisor, Charalampos Andriotis, for his continuous support and guidance throughout my work. From the onset, I was motivated and enthusiastic about this project, and my supervisor's feedback has only further fueled my passion for the subject matter. His feedback can be likened to a reward signal for a Deep Reinforcement Learning (DRL) agent, providing me with valuable information about the quality of my decisions and behaviors. Positive feedback has reinforced my approach and boosted my confidence, while negative feedback has challenged me to improve and strive for excellence. The feedback has allowed me to update my approach and make better decisions in the future. Thank you for your unwavering support, encouragement, and mentorship.

I want to thank my chair, Marc Ottele. Without your support and control over the process, I would still be working on the project and have lost my motivation a long time ago. I would also like to express my gratitude toward Stef Lhermitte. Thanks for being a source of inspiration in A.I. and climate change and being the sanity-check when I was too deep in the material.

This thesis and my entire study could not have been completed without the continuous support of my friends and family. I want to take this opportunity to shine a light on a few of them. Thank you, Ziead, for the help regarding the subject and for always being available for a helping hand. Thank you, Lars, for joining me on this journey, both in my previous study and Delft; thanks for sticking by my side and becoming a close friend. A special thanks go out to Baeda, my loving partner in this journey; you always listened to my struggles and supported me with your loving and continuous words of motivation. And last but not least, thank you to my parents, who always supported and believed in me in every way possible and made this journey possible.

Sammie Knoppert
Amsterdam, May 2023

Summary

In the last decades, climate change is causing our environment to change rapidly, unprecedented in recent history. Civil engineering structures are dependent on the deteriorating environment they are situated in. Changes can cause an increase in loading due to, for example, extreme weather events or alter the structure's resistance by, for instance, accelerated corrosion or an increase in the number of frost days. However, planning for such events depends significantly on the state of climate change, which is not considered in the sequential decision-making optimization for inspecting and maintaining our infrastructures.

Sequential decision-making optimization refers to the process of finding an optimal sequence of actions to be performed to keep the structure safe. Optimality often refers to the plan with the lowest costs. Traditional inspection and maintenance plans seek a solution by applying heuristic-decision rules on, for instance, time- or condition constraints and fail to find an optimal solution. During the last decade, a new method called Deep Reinforcement Learning (DRL) has been applied and proven to find an optimal strategy to beat traditional approaches. Especially the ability of DRL to find an optimal solution for partially observable environments makes it a perfect candidate for the problem at hand.

This thesis has developed a framework that incorporates partial observability over possible climate scenarios in the decision-making of engineering structures' inspection and maintenance planning. The framework explains the steps required to translate a physical system towards a Partially Observable Markov Decision Process (POMDP), as the mathematical framework needed to seek an optimal sequence of inspection and maintenance actions for. Then, it is explained how the POMDP can be used to find an optimal policy with a DRL algorithm, which includes benchmarking and testing the policy. A belief state has been incorporated over the climate scenarios to simulate the partial observability of the climate. Updating over the scenarios is provided by Bayesian Inference, using a climate parameter, such as temperature.

The framework has been applied to a case study in the second part of the thesis. The case study configures a stochastic deterioration process for different climate scenarios. An optimal policy has been found by applying Proximal Policy Optimization (PPO) with a decentralized policy for the various components. Two well-established heuristic-based maintenance policies, time-based maintenance (TBM) and condition-based maintenance (CBM) have been configured as benchmarks for the framework. The framework is compared against the benchmarks using lifecycle costs and safety as metrics. The policy provided by the framework has outperformed both benchmarks in terms of costs while maintaining the same safety. The policy beat TBM with 5% and CBM with 1% in terms of lifecycle costs. Another important distinction is that the benchmarks are optimized for the climate scenarios, while the framework finds this distinction without prior knowledge. It is therefore concluded that the framework can find an optimal policy under the uncertainties related to climate change. The case study, however, does not fully capture the complexity of engineering structures that the framework can catch. It is therefore recommended to use the framework for a more complex structure in the future.

Contents

List of Figures

List of Tables

1

Introduction

1.1. Problem description

1.1.1. Current practice

The costs to efficiently manage the safety of infrastructures takes up a significant amount of a country's budget. In a recent study done by the TNO [1], it is estimated that the projected costs to maintain infrastructure will continue to rise due to the deterioration of the structure and an increase in loading situations in the Netherlands. Currently, the Netherlands spends one billion euros annually to retrofit its infrastructure. This is projected to increase to three to four billion euros in 2040-2050 and even up to seven billion euros near the end of the century. With the annual cost of corrosion worldwide estimated to exceed $1.8 trillion, which translates to 3% to 4% of the Gross Domestic Product (GDP) of industrialized countries [2], one can imagine the impact of a significant increase in costs.

Due to climate change, the environment of our infrastructure is changing. A study by Nasr et al. [3] has found climate change to affect the state of our infrastructure in numerous risk categories, which can be grouped in the categories: durability, serviceability, geotechnical, increased demand, accidental loads, extreme natural events, and operational risks. Planning for such events, however, is difficult since we do not know what the state of the climate will be in the future. Typically, the potential impacts of climate change are projected in terms of climate scenarios to represent possible futures.

The environment of reinforced concrete (RC) infrastructure is changing. However, the possibility of a changing climate has not been considered during the design of existing RC bridges. Especially changes in temperature, relative humidity, and CO_2 concentration can significantly impact the deterioration of concrete bridges, as has been shown by various research. Bastidas-Arteage et al. [4] found, for instance, that climate change could reduce the time to failure of infrastructure due to chloride-induced corrosion by up to 31%. The department of climate change and energy efficiency of the Australian Government has done extensive research on the effects of climate change impacts on the deterioration of concrete structures ([5], [6], [7]) and made the following observations. Due to climate change, existing structures are more likely to suffer from decreased durability. However, the impact is difficult to assess for individual structures due to the widely varying locations, design, and environmental exposure. It is recommended to increase the amount of inspection and maintenance and develop adaptation options on an individual level, but could this be solved by implementing dynamic planning which considers these factors?

Traditional inspection and maintenance policies are based on heuristic decision rules, such as condition- or time-based maintenance. Condition-based policies focus on limits on the state of the environment, for instance, the amount of cracking in the concrete, and have corrective maintenance actions accordingly. Time-based policies require certain maintenance actions to be performed under specific periods, such as repairing the asphalt cover every five years. An optimal policy, however, would consider all the complexities in its decision-making, for example, the types of components, inspection- and maintenance actions, possible damage states, and the state of climate change for this study. However, the

solution space quickly becomes too large to explore (known as the curse of dimensionality). Therefore, it often resorts to heuristic-based approaches.

Deep Reinforcement Learning algorithms show a way to alleviate this curse. In recent years, research has been done on Deep Reinforcement Learning (DRL) inspection and maintenance planning for deteriorating structures. These algorithms are based on a branch within machine learning that employs an ”agent” that explores its environment and creates a dynamic policy based on its experiences, eventually converging towards an optimal policy. To use a deep reinforcement learning algorithm, the problem must be modeled as a Partially Observable Markov Decision Process (POMDP), which acts as the mathematical framework of the sequential decision-making process.

The agent employs a belief in the state of the environment at each decision step. The belief can be seen as a probability distribution over all possible environment states. By performing an inspection, the agent updates its belief and, therefore, gains more confidence about the state of the environment. At each time step, the agent decides to either do an inspection and/or do a maintenance action. After performing the action, the agent receives a reward. By simulating many lifetimes (episodes), the agent learns to maximize its reward function and find its optimal policy without prior knowledge of the environment. The policy found by a deep reinforcement learning approach can outperform every baseline policy, such as time- or condition based. The overall life-cycle reduction costs are between 5.8% and 22.3% [8].

1.1.2. Problem statement

Climate change increases the possibility of unsafe structures and, therefore, higher life cycle costs. With the costs of inspection and maintenance of infrastructure rising in the coming years, using more sophisticated policies is in dire need. It has been shown that individual maintenance plans are needed for existing infrastructure due to their varying nature in the spatial domain and design [5]. Dynamic policies learned by DRL algorithms can work on an individual level and take into account the added complexity with great success.

The scope of the thesis is centered around creating a new inspection and maintenance planning framework that incorporates the state of the climate within its decision-making. The framework exploits the possibility of DRL algorithms to find an optimal policy within a partially observable environment by employing a ’belief-state’ over the possible climate scenarios. The belief state is updated using Bayesian Updating, which uses data from a climate parameter as evidence for a scenario. The framework allows for planning for different future scenarios and acting accordingly. The coupling of the uncertainties related to possible climate change scenarios with a deteriorating structure into an inspection and maintenance plan for infrastructure has not yet been performed, which brings the novelty of the proposed framework.

The effects of climate change on the deteriorating structure and the modeling of a dynamic policy come with many assumptions and simplifications to remain within the scope. This raises the question of whether such a model can accurately betray a real-world structure. However, the framework’s power is that every single component can be extended to represent the real world more accurately in future research. The framework’s components are the climate, the structure, and the AI (DRL algorithm), as shown in figure 1.1.

1.2. Research objective

1.2.1. Research questions

With maintenance costs projected to increase in the coming years and climate change causing more deterioration of our infrastructure, more sophisticated maintenance programs will be needed to alleviate the costs. Recent studies show the potential of individual dynamic maintenance policies based on deep reinforcement learning to do so. This thesis' focus is on the development of a framework to incorporate climate change in decision-making, which leads to the following primary research question:

> *What is the effect of a framework that incorporates model uncertainty in the deteriorating environment of reinforced concrete due to climate change in the inspection and maintenance plan of bridges to lower the life-cycle costs and maintain safety?*

The main research question is broken down into the following sub-questions:

- How does a traditional inspection and maintenance plan compare to the dynamic policy from the proposed framework in a changing climate?
- What changes in the inspection and maintenance actions are proposed by the dynamic policy under a changing climate, and are they significant?

1.2.2. Objectives

To establish a scope for the research and create a clear view of the objectives to be accomplished, the following research objectives have been set up:

1. To model the effects of climate change on the deterioration of reinforced concrete.
2. To determine the effects of different climate scenarios on the deterioration rate of reinforced concrete structures.
3. To model a dynamic policy with the help of deep reinforcement learning, considering a changing climate.
4. To set up heuristic-based benchmarks to compare traditional methods against a dynamic policy and verify the results.
5. To compare the effects of different climate projections on RC structures' inspection and maintenance plan.

1.3. Methodology

The research aims to establish a framework to connect a changing climate, deterioration of reinforced concrete, and sequential decision-making in an inspection and maintenance plan. A dynamic policy adopts the inspection and maintenance plan on the fly based on observations of the climate and the structure. Such a policy will be compared with traditional policies to assess the need for more sophisticated methods due to changes in the deteriorating environment. The three main components are structure, climate, and optimal planning. Deep reinforcement learning is used to find the optimal sequence of inspection and maintenance actions based on information from the climate scenarios and the structure's condition. The framework with the links between the different components can be seen in figure 1.1, where AI represents the DRL algorithm.

A workflow has been adopted to find answers for the different research objectives, as explained below. Figure 1.2 gives an overview of the thesis structure and research outline. After the framework has been set up, it will be applied to an illustrative case study. The case study analyzes the results and answers the research questions.

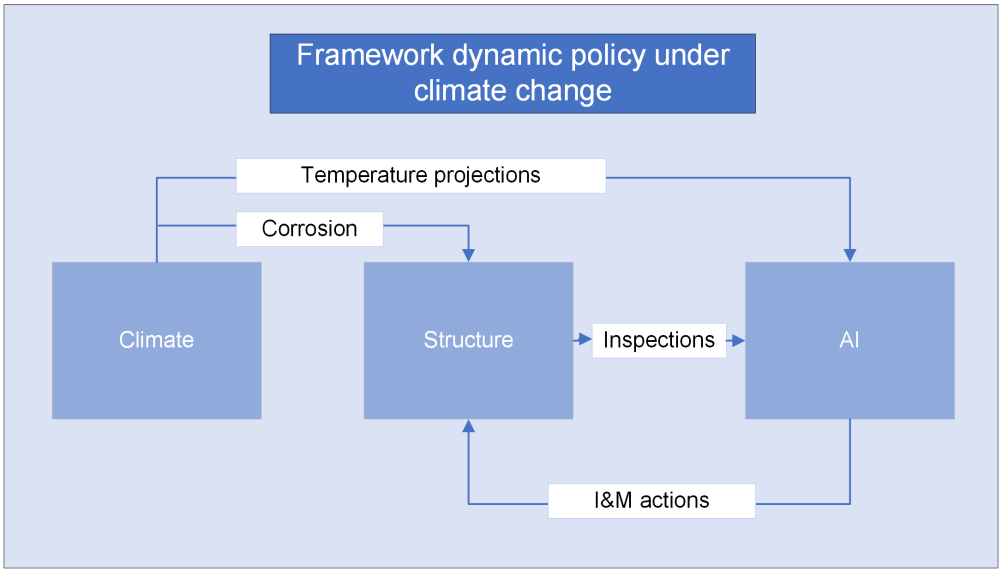


Figure 1.1: Framework of the dynamic policy under climate change

1.3.1. Data

The data used in this thesis is of a quantitative and secondary nature. Stochastic distributions of the involved parameters will be gathered from previously performed studies in the literature review. Statistical analysis methods have been used to analyze, validate, verify the data, and make conclusions accordingly.

1.3.2. Research outline

The thesis explores the available literature on the three main topics in chapter 2. The three main topics are the climate, stochastic deterioration of concrete, and optimal inspection and maintenance policies. The chapter is finalized with a summary of the current state of research and a closing paragraph that places the thesis in the existing literature. In chapter 3, the framework is developed and shows the requirements for each step. Next, the framework is applied to a case study in the second part of the thesis. First, the case study is introduced in chapter 4, and the physical system is described. Next, the case study is analyzed in chapter 5, and each step of the framework is performed. In this chapter, the first results will also be discussed. Chapter 6 covers the framework discussion and answers the research questions in the conclusion. A comprehensive overview of possible future works is also provided. The research outline is provided in figure 1.2.

Literature Review The literature review summarizes the theory and state of current research for the following topics: the effect of climate change on infrastructure, the deterioration of reinforced concrete, and deep reinforcement learning. The first part explains what climate change is, how climate scenarios are used to estimate the effect of global warming in the future, and what the effects of climate change are on infrastructures. The second part discusses concrete deterioration, where chloride-induced corrosion is taken as the governing deterioration mechanism. This is followed by the different models applied to model corrosion and how climate change affects corrosion. The third section covers the finding of the optimal I&M policy. In the introduction, traditional approaches will be explained. This is followed by an explanation of (PO)MDP, which gradually transitions to DRL algorithms.

Strategy of the framework The strategy of the framework applies the information gathered from the literature review to formulate the framework. The framework incorporates climate scenarios and stochastic deterioration of the different components to find the optimal sequence of inspection and maintenance actions. The framework is split into three main steps: the setup of the physical model, the setup of the POMDP, and finding the optimal policy.

Description of the case study The description of the case study describes the illustrative case study and applies this to the setup of the physical model, the first step of the framework. The result of this chapter is the required models and data of the physical system to represent the case study as a POMDP.

Analysis The analysis chapter covers applying the case study to the framework. It presents the results and corresponds to the second and third step of the framework. The results of this chapter are used to discuss the framework and answer the research questions. Also, an extra analysis is provided, which illustratively enhances the effects of climate change to make conclusions on the framework's effectiveness.

Discussion In the discussion, the results of the framework are discussed, conclusions are drawn, and the limitations are given. Also, a comprehensive review of possible future works is provided.

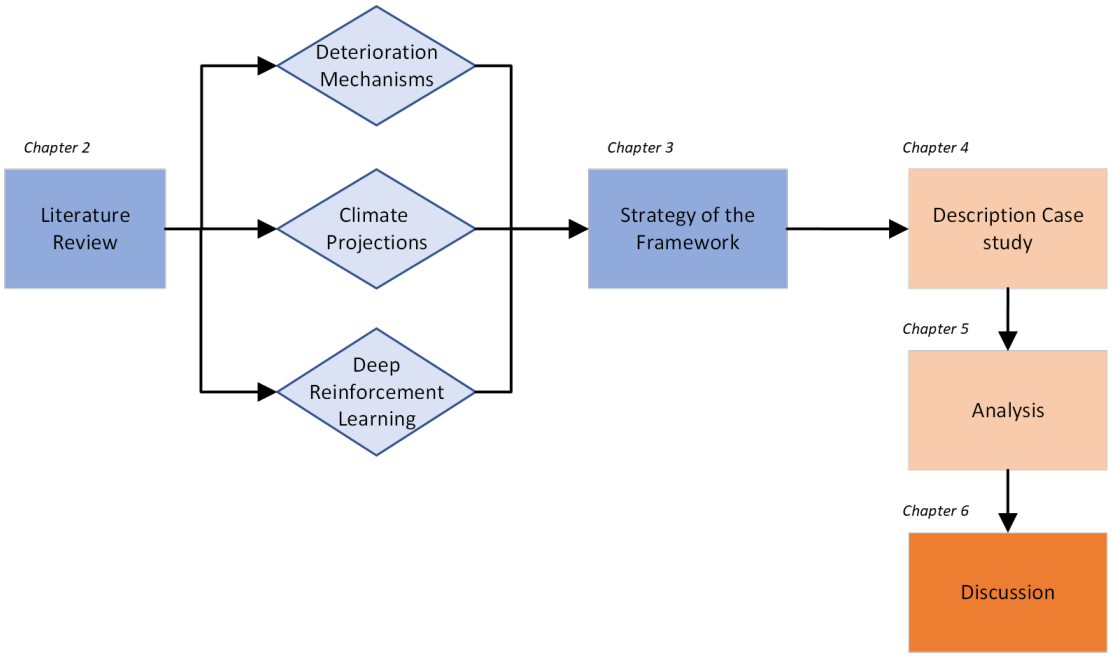


Figure 1.2: Thesis structure and research outline

Part I

Theory

2

Literature review

2.1. Climate change

The climate describes the average weather over an area for a prolonged period of time. Unlike the weather, the climate is predictable and can be seen as the statistics of the weather. Climate change is the change of the climate statistics that persist over an extended period of time [9].

In the recent decade, it has become common knowledge that climate change is occurring due to anthropogenic effects, especially due to the production of greenhouse gasses. The Intergovernmental Panel on Climate Change (IPCC) is the committee of the United Nations for assessing the science of climate change, identifying potential risks, and finding potential adaptation and mitigation strategies. At the moment of writing, the latest assessment report makes the following statements: Human influence is undoubtedly the cause of the warming of the oceans, atmosphere, and land and the increases in well-mixed greenhouse gas (GHG) concentrations since around 1750. Due to anthropogenic effects, the global surface temperature has evidently increased $0.8°C$ to $1.3°C$ from the second half of the nineteenth century to 2010-2019. Compared to the reference period global surface temperatures will very likely have increased to 1.0$°C$ to 1.8$°C$ for low emission scenarios, 2.1$°C$ to 3.5$°C$ for intermediate emission scenarios, and by 3.3$°C$ to 5.7$°C$ under high emission scenarios [10].

2.1.1. Climate scenarios

To analyze the effects of climate change and identify the possible future risks, the IPCC has created illustrative emission scenarios to set up climate scenarios. Socio-economic models are used to link changes in population, industrial development, and economics to emission scenarios. Illustrative emission scenarios are used to link changes in GHG emissions, air pollution, and land use. The combination of both is used to create climate scenarios to make predictions about future climate change scenarios [11].

The fifth assessment report [12] uses four Representative Concentration Pathways (RCPs) of GHG emissions, atmospheric concentrations, air pollutant emissions, and land use. The RCPs are used to assess the damages associated with different emission scenarios with the name of a pathway corresponding to the amount of radiative forcing by the end of the century [W/m^2]. The following RCPs are used in AR5 (see figure 2.1 [12]).

- RCP2.6 - strict mitigation, highest peak emission in 2020, negative emissions at the end of the century.
- RCP4.5 - intermediate scenario, Peak emissions around 2040 followed by a decline in emissions, stabilizing around the year 2080.
- RCP6.0 - intermediate scenario, Peak emissions around 2080, followed by a decline in emissions until the end of the century.
- RCP8.5 - no mitigation, business as usual.

Climate models are used as a mathematical representation of processes involved in the climate system of the Earth. A variety of models are used to assess the effects of climate change, from simple idealized models to general circulation models (GCMs) [12]. GCMs are numerical models to represent the physical processes of the atmosphere, ocean, cryosphere, and land surface. They are used in order to forecast the weather, understand different aspects of the climate, and forecast climate change using the RCPs given above [9].

The Royal Netherlands Meteorological Institute (KNMI) produces climate scenarios for the Netherlands based on the assessment reports published by the IPCC. The latest national climate scenarios are based on the fifth assessment report of the IPCC and are explained below [13].
The KNMI has set up national climate scenarios using the calculations performed by the IPCC with observational evidence, augmented with the climate model of the KNMI for Europe [13] (Figure 2.2). Each climate projection of the KNMI is based on two variables, the amount of worldwide temperature increase following an RCP pathway (G = moderate, W = warm) and changes in airflow patterns (see figure 2.2a). The worldwide temperature increase is based on two emission scenarios: RCP4.5 (stabilization) and RCP8.5 (high emissions) from the IPCC [12] (see figure 2.2b). The changes in airflow patterns correspond to a low influence of circulation change (L) and a high influence of circulation change (H).
Each scenario is depicted by key values about the change in twelve climate variables, including temperature, precipitation, sea level, and wind. Key values are used to describe the climate for the different climate scenarios (GL, GW, WL, WH) and are given based on three-time frames: reference data (1981 - 2010), mid-century (2050), and end of the century (2085). The projected key values for the mid- and end century also come with a 30-year natural variability to provide a 90 % confidence range, meaning that there is a 10% probability that the 'real' climate deviates from the bandwidth of the climate scenarios (the colored band in figure 2.2b represents the confidence range) [14].

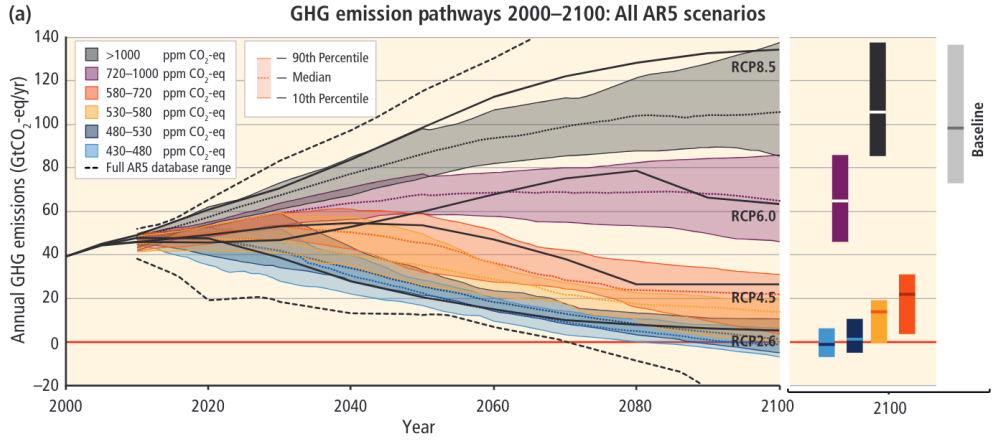
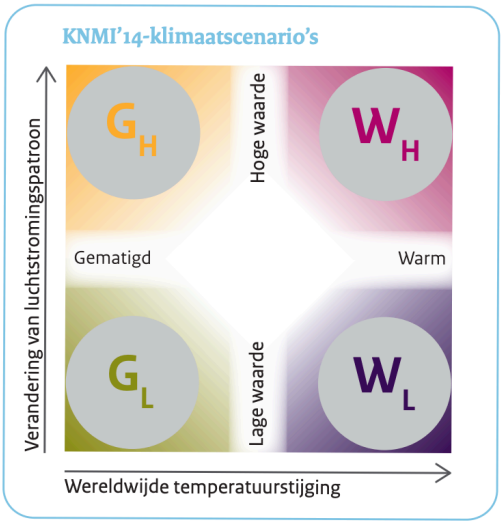
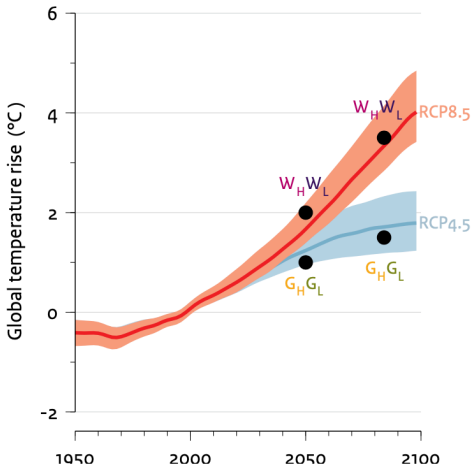


Figure 2.1: GHG emission pathways for the different RCPs [12]



(a) Construction of climate scenarios from KNMI based on global temperature rise (horizontal axis) and changes in airflow patterns (vertical axis)



(b) Global temperature rise based on climate model calculations performed by the IPCC [12] and linked with the dutch climate scenarios

Figure 2.2: Climate scenarios from the Royal Netherlands Meteorological institute [13]

2.1.2. Effect on infrastructure

Numerous studies have been performed on the link between climate change and the effects on infrastructure with Nasr et al. giving an extensive review of the effects [3]. The climate change imposed risks are divided into the following categories: durability, serviceability, geotechnical risks, increased demand, accidental loads, extreme natural events, and operational risks. Combinations of the risk categories lead to reductions in safety and increases in costs. The study presents that two climatic parameters may cause the highest risk increase, higher future temperatures, and an increase in precipitation in some regions [3].

Climate change alters the aggressive environment, mainly in CO_2 concentrations, humidity, and temperature. These changes can accelerate the deterioration process which leads to a decline in the safety, serviceability, and durability of RC infrastructure [15]. With the annual costs of corrosion worldwide to exceed 1.8 trillion dollars, which is 3-4 % of the Gross-Domestic Product (GDP) of industrialized countries [2], one can imagine that an accelerated corrosive environment due to climate change can result in significant increases in damages and costs.

Corrosion is highly dependent on climate change, with the corrosion initiation and propagation rate being affected directly by humidity and temperature [5]. The effects of accelerated degradation on infrastructure have been studied in numerous studies ([5], [6], [7], [4], [16], [17], [15], [18]). Nasr et al, found for instance that the probability of failure at the end of the century can be considerable, with a $\pm 123\%$ increase due to chloride-induced deterioration [17]. Bastidas-Arteaga et al, found that the effects of global warming could reduce the time to failure due to chloride-induced corrosion by up to 31% [4].

The accelerated degradation of reinforced concrete structures combined with the high costs of corrosion requires a new approach to maintenance. The implementation of the effects of climate change on existing concrete structures is vital for inspection and maintenance planning [5].

2.2. Deterioration of reinforced concrete

The deterioration of reinforced concrete (RC) infrastructure occurs due to degradation of the concrete or corrosion of the embedded reinforcement. Degradation of the concrete occurs due to sulfate attacks, Alkali-Silica reactions (ASR), freeze-thaw cycles, and fatigue. The reinforcement degrades due

to corrosion and is caused by two major sources: carbonation-induced corrosion and chloride-induced corrosion [19].

Corrosion is generally of most concern since the reduction in the cross-sectional area will in time lead to a loss in strength of the reinforcement and cracking or spalling of the concrete [19]. The annual costs of corrosion annually are estimated to exceed $\$1.8$ trillion worldwide, which translates to 3% to 4% of the gross domestic product of developed countries [2]. The development of more sophisticated inspection and maintenance planning for infrastructures subjected to corrosion can therefore have a big impact on the overall life-cycle costs.

The steel bars in the concrete are protected by a thin passive layer of oxides, protecting them from oxygen and water, which in turn can lead to corrosion and the production of rust. The corrosion leads to a decrease in the cross-sectional area of the steel bar, which in turn leads to a reduction in strength. The rust products produced by the corrosion process have a higher volume than the original steel bar which causes expansive stresses to occur. The stresses cause the concrete to crack, delaminate, and eventually spall. The two major sources of the removal of the passive layer are carbonation-induced corrosion and chloride-induced corrosion [20].

2.2.1. Carbonation-induced corrosion

The CO_2 in the air reacts with the hydrated cement in the presence of moisture. The hydrate in the cement which reacts most promptly with CO_2 is $Ca(OH)_2$, which produces $CaCO_3$ known as calcium carbonate. The importance of carbonation is that the pH of the pore water is reduced from 12.6-13.5 to about 9. Once the reduction of the pH reaches the vicinity of the steel bars, the protective oxide layer is destroyed. This leads in turn to corrosion to occur, given that moisture and oxygen are present [20].

Although several studies have shown that climate change can accelerate carbonation-induced corrosion ([5], [6], [7], [4], [18], [21]), chloride-induced corrosion is found to be most costly to repair [22] and to be the governing deteriorating mechanism for structures subjected to deicing salts in the winter and marine environments [19].
The combined effects of chloride-induced corrosion and carbonation-induced corrosion are not well studied and no quantitative models exist [6].

Since chloride-induced corrosion is found to be the most costly and governing deterioration mechanism for structures subjected to deicing salts and no quantitative models exist to combine the effects of chloride-induced corrosion and carbonation-induced corrosion, only chloride-induced corrosion will be considered as the deteriorating mechanism in this thesis.

2.2.2. Chloride-induced corrosion

Chloride-ions are present at the infrastructure due to de-icing salts, contact with seawater, or deposited by air-borne droplets of seawater. The chloride ions penetrate the concrete by transport of water containing chlorides and diffusion of the ions. Once chloride ions penetrate the concrete cover and accumulate to a critical level around the rebar, the passive layer is removed which will result in the start of corrosion [20].

The deterioration of the embedded reinforcement is typically described by two phases: (i) The initiation phase, the time required for the deleterious substance to reach the reinforcement and depassivate the protective layer around the rebar; (ii) The propagation phase, the actual corrosion of the rebar [23].

Corrosion initiation

The time required for chloride-ions to penetrate the concrete by diffusion and remove the passive layer is typically described by the model developed by Duracrete [24] based on Fick's second law of diffusion, see equation 2.1. Once the chloride concentration $C(x,t)$ is equal to the critical chloride concentration C_{crit} at the height of the rebar ($x = cover$), corrosion is assumed to have initiated.

$$C(x,t) = C_0 \left[1 - erf \left(\frac{x}{2\sqrt{k_e k_t k_c D_c (\frac{t_0}{t})^{n_d}}} \right) \right] \tag{2.1}$$

where: $C(x,t)$ is the chloride content at a distance x from the surface at time t, C_0 is the surface chloride concentration, D_c is the chloride diffusion coefficient, erf is the error function; k_e, k_t, k_c are environmental, testing and curing method factors, t_0 is the reference period, t is the time and n is the aging factor of the concrete.

However, it has been shown that the predicted time for corrosion initiation is highly variable especially when input is derived from engineering structures instead of laboratory samples. This poor predictability is mainly caused by three reasons: a lack of understanding of corrosion initiation, the use of non-representative laboratory samples, and neglect of the size effects for localized corrosion [25].
According to a probabilistic evaluation, the probability of corrosion initiation after fifty years is ±60% for a concrete wall and ±90% for a concrete column, examined in the study performed by Val et al, [26].
The Netherlands built most of its existing reinforced concrete infrastructure in the period 1960-1980, having a lifespan of around 60 years [27].

The lack of thorough understanding of the corrosion initiation time combined with the high likelihood that corrosion has already been initiated leads to the assumption that the corrosion process has already started for the remainder of this thesis.

Corrosion rate

Several models exist to predict the corrosion rate once corrosion has initiated with Lu et al. [28] giving an overview of the models. Many of the more advanced models require data that is unknown for infrastructures. Therefore in this research, the model applied by DuraCrete [24] will be used which uses a time-invariant model but temperature dependent (see equation 2.2), due to a lack of data on the time-dependent effects on the corrosion rate [5]. The same model is applied by Wang et al. [6] and Bastidas-Arteaga et al. [4].

$$i_{corr}(t) = i_{corr,20} \cdot [1 + K_T(T(t) - 20)] \tag{2.2}$$

where: $i_{corr,20}$ is the reference corrosion current density at $T = 20°C$ and follows from Duracrete [24], K_T is a factor depending on the value of T :($K_T = 0.025$ if $T(t) < 20°C$ or $K_t = 0.073$ if $T(t) > 20°C$), and $T(t)$ being the annual average temperature ($°C$) at time t.

According to DuraCrete [24], structures subjected to deicing salts need to have variables depending on the material or environmental properties to be chosen similarly to structures in a marine environment/splash zone.

Studies have shown that humidity can have a significant impact on the corrosion rate. Corrosion becomes negligible once the relative humidity drops below 50 % or once the concrete is fully submerged. The optimal range of relative humidity is between 70% and 80 % [20].

There is a lack of data for the climate scenarios regarding relative humidity for different climate scenarios in the Netherlands and a small change is expected due to climate change in the relative humidity [13]. The relative humidity in the Netherlands is typically within the optimal range for corrosion [29]. The combination of lack of data, small expected changes, and the annual humidity in the Netherlands being in the optimal range lead to the assumption that relative humidity will not be implemented in the modeling of corrosion for the remainder of this research.

Chloride-induced corrosion propagation

The modeling of corrosion propagation in the presence of climate change is typically done based on four limit states: the probability of corrosion initiation, crack initiation, spalling, and failure ([4], [6], [17], [16]). In this thesis, the focus will be on spalling of the concrete and the failure of the structure. Crack initiation is taken into account in the time required for spalling to occur.

The loss of reinforcement steel by chloride-induced corrosion can be divided into two types of corrosion, uniform corrosion, and pitting corrosion ([23], [19], [30]). The combination of both uniform and pitting corrosion is described by Val and Melchers [31].

1. *Uniform corrosion*

   Uniform corrosion is described by a uniform loss of the diameter of the rebar. The time-dependent uniform cross-sectional area of a reinforcement bar ($A_{st,U}(t)$) can be described as follows [23]:

$$A_{st,U}(t) = \begin{cases} d_{r,0}^2 \cdot \frac{\pi}{4} & \text{if } t \leq t_i \\ d_r(t)^2 \cdot \frac{\pi}{4} & \text{if } t_i < t < t_i + \frac{d_{r,0}}{r_{corr}} \\ 0 & \text{if } t \geq t_i + \frac{d_{r,0}}{r_{corr}} \end{cases} \tag{2.3}$$

$$d_r(t) = d_{r,0} - 0.0232 \cdot \int_{t_i}^{t} i_{corr}(t)dt \tag{2.4}$$

$$r_{corr}(t) = 0.0116 \cdot i_{corr}(t) \tag{2.5}$$

   where: $d_{r,0}$ is the initial reinforcement diameter, $d_r(t)$ is the current reinforcement diameter at time t, r_{corr} is the annual metal loss per unit surface area [$\frac{mm}{year}$] and t_i being the initiation time of the corrosion.

2. *Pitting corrosion*

   The model applied for pitting corrosion is the hemispherical model derived by Val and Melchers [31]. The model describes the residual cross-section due to pitting corrosion $A_{st,P}(t)$ for each timestep t.

$$A_{st,P}(t) = \begin{cases} A_{st,0} - A_1 - A_2 & \text{if } p(t) \leq \frac{d_{r,0}}{\sqrt{2}} \\ A_1 - A_2 & \text{if } \frac{d_{r,0}}{\sqrt{2}} < p(t) \leq d_{r,0} \\ 0 & p(t) > d_{r,0} \end{cases} \tag{2.6}$$

   with:

$$A_1 = \frac{1}{2}\left(\theta_1 \left(\frac{d_{r,0}}{2}\right)^2 - a\left|\frac{d_{r,0}}{2} - \frac{p(t)^2}{d_{r,0}}\right|\right) \tag{2.7}$$

$$A_2 = \frac{1}{2}\left(\theta_2 p(t)^2 - a\frac{p(t)^2}{d_{r,0}}\right) \tag{2.8}$$

$$\theta_1 = 2\arcsin\left(\frac{a}{d_{r,0}}\right) \tag{2.9}$$

$$\theta_2 = 2\arcsin\left(\frac{a}{2p(t)}\right) \tag{2.10}$$

$$a = 2p(t)\sqrt{1 - \left(\frac{p(t)}{d_{r,0}}\right)^2} \tag{2.11}$$

$$p(t) = 0.0116 \cdot i_{corr}(t)R \tag{2.12}$$

   where: $A_{st,0}$ is the original reinforcement area before deterioration, $p(t)$ is the pitting penetration, R is the ratio of pitting penetration to uniform corrosion taken as a uniform distribution between 4-8 [31].

3. *Residual cross-sectional area*

   The combination of uniform corrosion $A_{st,U}(t)$ and pitting corrosion $A_{st,P}$ can be computed as follows [17]:

$$A_{st} = (A_{st,U}(t) - A_{st,0})\left(1 - \frac{a}{2d_{r,0}}\right) + A_{st,P}(t) \tag{2.13}$$

4. *Concrete spalling*

During the process of corrosion, corrosion products are produced which have a higher volume than the original rebar. These products cause internal stresses to occur within the concrete. The stresses cause the concrete to crack, delaminate and eventually spall [20].

The process of spalling is typically described by the time required for initiation of cover cracking (t_{1st}) and the time to severe cracking (t_{sev}). Time to spalling is the summation of both terms: $t_{spal} = t_{1st} + t_{sev}$. Many empirical, numerical, and analytical models exist to predict the time to cover cracking [32]. It has been studied that a simpler model depending only on, for instance, the corrosion current density $i_{corr}(t)$ and the cover height generally perform better than more sophisticated methods [32]. This led to the choice to adopt the model proposed by Rodriguez et al, [33] and later employed by DuraCrete [24]:

$$x_{cr} = 83.8 + 7.4\left(\frac{c}{d_{r,0}}\right) - 22.6 f_{ct} \tag{2.14}$$

$$t_{1st} = \frac{x_{cr}}{11.6 \cdot i_{corr}} \tag{2.15}$$

where: x_{cr} is the critical corrosion penetration for cover cracking, c is the thickness of the concrete cover, f_{ct} is the splitting tensile strength of the concrete.

Similar to the time required for cracking initiation, many models exist to model severe cracking. In this study the model developed by Mullard and Stewart [34] will be used, in line with the following studies [6], [17]. The time required for severe cracking (t_{sev}) is described for a crack to propagate to a width of 1 mm under a constant corrosion rate [17].

$$t_{sev} = k_R\left(\frac{w - 0.05}{k_c ME(r_{crack}) r_{crack}}\right) \cdot \left(\frac{0.0114}{i_{corr-20}}\right) \tag{2.16}$$

with:

$$k_R \approx 0.95\left[exp\left(-\frac{0.3 i_{corr(exp)}}{i_{corr-20}}\right) - \frac{i_{corr(exp)}}{2500 i_{corr-20}} + 0.3\right] \tag{2.17}$$

$$r_{crack} = 0.0008 exp(-1.7\phi_{cp}), 0.1 \leq \phi_{cp} \leq 1.0 \tag{2.18}$$

$$\phi_{cp} = \frac{c}{d_{r,0} f_t} \tag{2.19}$$

where: k_R is a correction factor for the rate of loading ($i_{corr(exp)} = 100\mu A/cm^2$), k_c is a confinement factor , w is the limit crack width ($1mm$), $ME(r_{crack})$ is a model error represented by a normal distribution $N(\mu = 1.04, COV = 0.09)$, r_{crack} is the rate of corrosion propagation, f_t is the tensile strength of concrete.

In order to incorporate the temperature-dependent corrosion rate, the approach taken by [6] is used. The amount of corrosion product (m_{corr}) under a constant corrosion rate ($i_{corr-20}$) is calculated by combining the equations 2.15 and 2.16:

$$m_{corr} = \frac{x_{cr}}{11.6} + 0.0114 k_R\left(\frac{w - 0.05}{k_c ME(r_{crack}) r_{crack}}\right) \tag{2.20}$$

Then the translation is made from the corrosion product under a constant corrosion rate to the temperature-dependent corrosion rate:

$$m_{corr} = \int_{t_0}^{t_{sp}} i_{corr}(t) dt \tag{2.21}$$

By combining equation 2.20 and 2.21, a relationship is found between the corrosion rate and the amount of corrosion product required for spalling. The relationship can be used to obtain projections of the time to spalling for different climate scenarios. By predicting the effects of a changing temperature, one can adopt different inspection and maintenance plans accordingly.

2.2.3. Deterioration modelling

The model used to describe the cross-sectional damage due to chloride-induced corrosion in section 2.2.2 is known as a Random-Variable Degradation (RVD) model, where the average rate of deterioration has a probability distribution. An RVD model is unable to capture the temporal variability associated with the evolution of corrosion damage. In order to model the temporal variability, a Stochastic-Process Deterioration (SPD) model as thoroughly explained in [35] and [36] needs to be adopted. An important aspect of the modeling of continuous deterioration is that there cannot be negative increments of deterioration, eliminating the use of a Brownian motion with drift model. The gamma process is an SPD model with independent, positive increments following a gamma distribution characterized by an identical scale parameter and a time-dependent shape parameter. The application of the gamma process for the maintenance of infrastructure has been shown to be a good modeling choice and suitable to model gradual damage accumulation such as corrosion [36].

If the damage $d = \frac{A_s(t)-A_s(0)}{A_s(0)}$ is a random variable that has a gamma distribution with a shape parameter $v(t) > 0$ and scale parameter $u > 0$, then the probability density function for each timestep t is described as:

$$Ga\left(d|v(t),u\right) = \frac{u^{v(t)}}{\Gamma(v(t))} \cdot exp(-ud) I_{(0,\infty)}(d \geq 0) \tag{2.22}$$

with,

$$\Gamma(v(t)) = \int_{z=0}^{\infty} z^{v(t)-1} exp(-v(t)) dz \tag{2.23}$$

and $I_{0,\infty}(d)$ being an indicator function to assure positivity in d. With $v(t)$ being the increasing, real-valued shape function for $t \geq 0$ with $v(0) = 0$ and a scale parameter with $u > 0$, then a continuous-time SPD model is found with the following properties:

1. $d(0) = 0$
2. $d(t+1) - d(t) \sim Ga(v(t+1) - v(t), u)$
3. $d(t)$ has independent increments

The mean value and standard deviation of the gamma distribution are described as:

$$d_m = \frac{v(t)}{u}, \sigma = \frac{\sqrt{v(t)}}{u} \tag{2.24}$$

The second property ensures that increments in damage also follow a gamma distribution. This property ensures that only positive, independent damage samples can occur.

The deterioration model as described in 2.2.2 equation 2.13 is proportional to a power-law function of the form $A_{st}(t) \propto A * t^{\beta}$, where different shape-functions can be used to model different deterioration rates as is expected for a changing climate.

A gamma process can be calibrated by an RVD model by matching the mean damage and standard deviation, using equation 2.24. An example of an arbitrary non-stationary gamma process modeling cross-sectional area damage is given in figure 2.3.

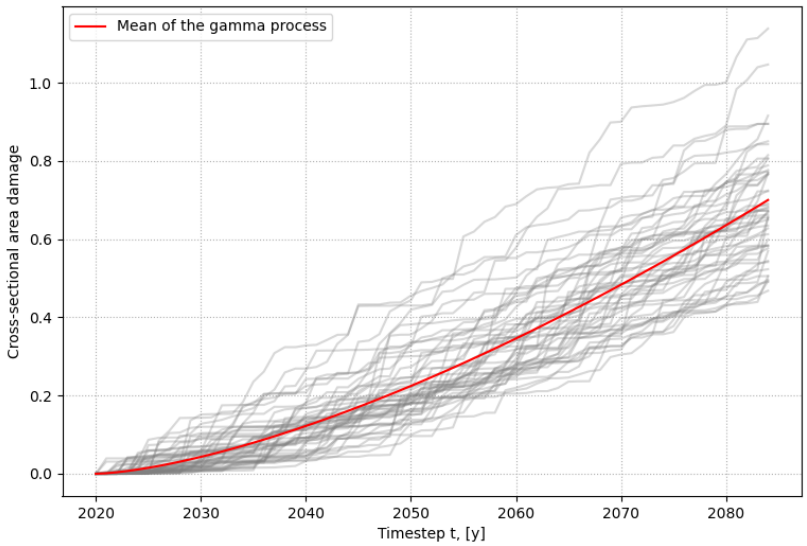


Figure 2.3: Random realizations of a Gamma Process describing cross-sectional damage

2.3. Optimal inspection and maintenance policies

Inspection and maintenance policies for civil infrastructures are dependent on heuristic decision rules. Heuristic-based policies apply time- or performance metrics to determine if an inspection or maintenance must be performed. Most of these policies are either time-based maintenance policies or condition-based maintenance policies [37]. Time-based maintenance policies fall under the category of preventive maintenance. In time-based maintenance, specific time intervals are used to assign maintenance actions for the structure such as replacing or repairing a component. Condition-based maintenance policies assign maintenance actions based on performance metrics, i.e. the amount of corrosion or crack width. The inspection and maintenance policy used by the Netherlands falls under the category of condition-based maintenance. According to the fact sheet of Rijkswaterstaat [38], three different types of inspections are performed on predetermined time intervals. The inspection reports are used as performance metrics on which the inspection and maintenance policies are grounded.

The optimality of such an inspection and maintenance policy based on heuristic rules depends strongly on the experience of the designer [39]. Condition-based and time-based maintenance policies use exhaustive policy search to solve the life-cycle problem which scales poorly for multiple-component systems or long planning horizons [8]. In recent years a new field of machine learning called deep reinforcement learning has been applied to the structural maintenance domain and has been found to outperform both condition-based and time-based maintenance policies for complex engineering systems. For instance, Andriotis et al [8], outperformed the optimal benchmarks for a complex engineering system with a reduction in life-cycle costs between 5.8% to 22.3%.

The focus of this section is the exploration of methods to find an optimal sequence of inspection and maintenance actions for a civil engineering structure under uncertain action outcomes and climate scenarios. First, an explanation will be given of the mathematical framework behind stochastic-control processes known as Markov Decision Processes (MDPs) with an extension to partially observable environments. Followed by solving methods to find an optimal policy under stochastic control.

2.3.1. Markov decision process

The optimal control of stochastic processes under exact observations and uncertain outcomes of different actions is typically described using Markov Decision Processes (MDPs). An MDP employs a 'decision-maker' who interacts with a stochastic environment by performing actions with uncertain outcomes at each discrete decision step.

The theoretical background on how MDPs work and extensions on the framework (except for Partially Observable MDPs) will not be discussed in this literature review due to the practical nature of this study and the existence of many reviews of the material in existing literature, for instance, [40] provides a good review within a structural health application. In this subsection, the basic concepts of the MDP and the interaction between the environment and the decision-maker will be introduced as background for later concepts of the framework introduced in this research.

At each discrete timestep t, an MDP assumes that the environment is in current state $s \in S$, takes an action $a \in A$, receives a reward based on the current state and chosen action $r = R(s_t, a_t) \in R$ and transitions towards the succeeding state $s' \in S$ according to a stochastic transition model $P(s'|s, a)$. It is assumed that the agent observes the succeeding state with full certainty. (In the formulation above: the state, action, and reward correspond to timestep t. The succeeding timestep $t+1$ is denoted by an apostrophe '.) Typically, a discount factor γ is also included to discount future rewards. An MDP can therefore be described according to the following tuple: $< S, A, R, P, \gamma >$.

Inherent to the MDP is the Markov property which states that the past of s is independent of its future and conditionally dependent on the present. The Markov Property states that the current state s provides all the required information together with the action a taken at decision step t to formulate the successor state s'. The interaction of the decision-maker or 'agent' can be seen in figure 2.4.

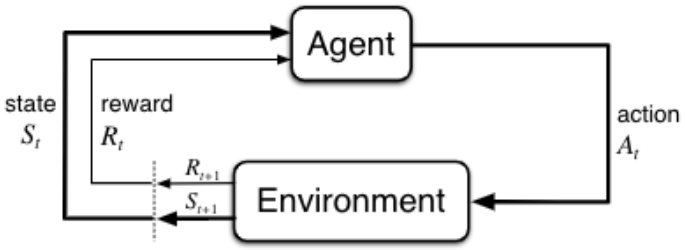


Figure 2.4: Agent-environment interaction [41]

At each time step, the decision-maker decides what action to take under the current observation of the environment. This process is called 'sequential decision making'. The sequential-decision making performed by the decision-maker or 'agent' is optimized by introducing an objective in the form of a long-term reward function. The long-term reward function is case specific but the most general form is the maximization of the expected total rewards over the lifetime. In a structural maintenance setting, life-cycle costs can be used. Costs can be seen as negative rewards so the maximization of total rewards can also be viewed as the minimization of life-cycle costs. Since this is a stochastic control problem, the expected total rewards will be used. Both the terms rewards and costs will be used simultaneously since one is just the inverse of the other. The objective is given below in equation 2.25.

$$\max \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t r_t\right] = \min \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t c_t\right] \tag{2.25}$$

where: $\mathbb{E}$ is an expectation operator, T is the planning horizon, γ denotes the discount factor $\gamma \in [0, 1]$, $c_t = -r_t$ is the cost received at time-step t.

The summation of total rewards is also known within the field of dynamic programming as the total return, G_t, defined below:

$$G_t = R_{t+1}(s_t, a_t) + \gamma R_{t+2}(s_{t+1}, a_{t+1}) + ... = \sum_{k=0}^{T} \gamma^k R_{t+k+1} \tag{2.26}$$

A policy is typically characterized by means of a value-function $v_\pi(s)$ or action-value function $q_\pi(s, a)$. The value functions are used to describe how good it is to be in a given state s (or state-action pair (s, a) for the action-value function). The value function is represented in equation 2.27 in terms of the total return.

$$v_\pi(s) = \mathbb{E}_\pi(G_t|s) \tag{2.27}$$

The action-value function is similar to the value function, however, also includes the action taken at decision step t and following the policy π thereafter.

$$q_\pi(s,a) = \mathbb{E}_\pi(G_t|s,a) \tag{2.28}$$

The value function can be seen as taking the expected value of the action values over all possible actions in the following manner:

$$v_\pi(s) = \mathbb{E}_{a\in A}[q_\pi(s,a)] \tag{2.29}$$

By decomposing the total return into the immediate return and the expected reward for the succeeding time steps in the action-value function, a recursive relationship is found which can be exploited to find the optimal policy.

$$q_\pi(s,a) = \mathbb{E}[R(s,a) + \gamma G_{t+1}|s,a] = R(s,a) + \mathbb{E}[\gamma q_\pi(s',a')|s,a] \tag{2.30}$$

The optimal sequence of actions found by solving the long-term goal is known as the optimal policy π^*. A policy π maps a state to an action to be performed. The optimal policy is typically non-stationary, meaning that different time steps lead to a different mapping of states to actions under a finite time horizon [40].

For finite MDPs, there exists a unique optimal policy π_* which maximizes the expected total return over its lifetime. The relation between the value of the current state and the value of its successor states is known as the Bellman equation. This equation can be exploited to break one big problem into many subproblems which will be used to find the optimal policy π_*[41]. An optimal policy has the maximum value/action value of all possible policies. Applying the maximization on the value function leads to the following Bellman optimality equation.

$$v_*(s) = \max_{a\in A}\left[R(s,a) + \gamma \sum_{s'\in S} P(s'|s,a)\cdot v(s')\right] \tag{2.31}$$

The Bellman optimality equation can also be rewritten using the action values and the relation with the value function given in equation 2.29 in the following manner.

$$q_*(s,a) = \left[R(s,a) + \gamma \sum_{s'\in S} P(s'|s,a)\cdot v(s')\right] \tag{2.32}$$

Once one has found v_* or q_*, it is easy to determine the optimal policy. For the optimal value function v_*, an algorithm simply has to perform a one-step look ahead to find the optimal action and act greedily. For the action-value function q_*, it is even more straightforward since it can just select the action with the highest action value. However, the trick is solving this optimality equation to find these values. The optimality equation is non-linear and requires high computational costs. This especially becomes a problem in the case of a real-world example where high dimensions of states and actions are common, making it impossible to calculate the value- (or action-value) function exactly and therefore requiring approximation methods [41].

Computation of the optimal policy is traditionally performed using algorithms that fall under the category of ”Dynamic Programming”. Dynamic programming is a method to solve MDPs given a perfect model of the environment. Such algorithms include policy iteration or value iteration. The problem of finding the optimal policy is prone to what is known as ”the curse of dimensionality”. The size of the problem grows exponentially with the number of states and actions [8]. In Dynamic Programming, the decision maker must compute and store values for each state-action pair, which becomes infeasible for real-world problems with big state and action spaces.

The requirement of offline knowledge of the perfect model of the environment and small dimensions in state and action spaces makes the use of dynamic programming methods absolute for complex civil engineering structures. More advanced techniques such as reinforcement learning and deep reinforcement learning algorithms which will be discussed in sections 2.3.2 and 2.3.3 are required.

Partially Observable Markov Decision Processes

In MDPs, a basic assumption is that at each decision step, before performing an action, a perfect inspection is performed. Inspections at each decision step, let alone perfect inspections, do not correspond to how an inspection and maintenance planning is performed for civil engineering structures. Within the framework of this research on corroding reinforced concrete structures, this assumption becomes even bolder since currently all the non-destructive inspection methods only approximate the current state [40].

Partially Observable Markov Decision Processes (POMDPs) are an extension on MDPs that give way for the possibility to include a belief **b** about the condition states $b(s)$ of the environment to give the possibility for imperfect inspections. The belief is a probability distribution over the states of the environment S.

The belief of the decision maker gets updated by performing an action a and making observation $o \in O$ over the new state s'. The action can represent an inspection or maintenance action. Updating of the belief is done according to Bayesian updates at each decision step [40].

$$b(s') = \frac{p(o|s', a)}{p(o|\mathbf{b}, a)} \sum_{s \in S} p(s'|s, a) b(s) \tag{2.33}$$

where: $b(s)$ is the probability of being in state s, when all states $s \in S$ are combined form the array **b**, $p(o|\mathbf{b}, a)$ is known as the normalizing constant, the sum of the likelihood ($p(o|s', a)$) and prior distribution $\left(\sum_{s \in S} p(s'|s, a) b(s)\right)$ product:

$$p(o|\mathbf{b}, a) = \sum_{s' \in S} p(o|s', a) \sum_{s \in S} p(s'|s, a) b(s) \tag{2.34}$$

An MDP was described by a tuple of $MDP =< S, A, R, P, \gamma >$, a POMDP is described by a tuple of $POMDP =< S, A, R, P, \gamma, O, Z >$, where O is a finite set of observations and Z is the observation model, modeling the effects of states and actions on observations $Z(s', a) = p(o|s', a)$.

Implementation of the belief state and observations in the optimal value function given in equation 2.31 of the MDP gives:

$$v_*(\mathbf{b}) = \max_{a \in A} \left[\sum_{s \in S} b(s) r(s, a) + \gamma \sum_{o \in O} p(o|\mathbf{b}, a) v(\mathbf{b}') \right] \tag{2.35}$$

The MDP and POMDP share many similarities which can be seen in the formulation of the environment and similarities in the definition of the optimality equation in 2.35. The inclusion of imperfect observation does come at a cost, POMDPs are notoriously harder to solve. A POMDP extends on the ”curse of dimensionality” by having a continuous belief-state space to plan for and in addition, is also subjected to ”the curse of history” [40]. The curse of history relates to the problem that the number of possible actions and observations grows exponentially with the planning horizon. Both problems cause an exponential increase in computational effort and major problems in the adaptation of POMDPs to large-scale problems.

In order to overcome the complexity of solving a POMDP, typically algorithms such as point-based solvers are used to approximate the value function, and an analysis of several point-based algorithms is performed by Andriotis et al. [42]. Applications of such point-based value iteration solvers to the structural maintenance domain can be found in the following studies: [40], [43], [44]. The problem with point-based solvers is that they cannot handle the high-dimensionality of state, action, and observation spaces which are typical for structural systems [39] [45].

2.3.2. Reinforcement learning

One of the problems encountered with dynamic programming algorithms, such as value-iteration or policy iteration, is the assumption of offline knowledge about the environment. Unlike the previous algorithms, reinforcement learning is able to acquire knowledge about the environment by interacting with it by performing actions. After performing an action and receiving a reward, the agent changes its policy accordingly to find an optimal policy in terms of its objective function. This makes the algorithm

more robust and applicable in many different use cases. It learns about the environment by 'exploration' and finds an optimal path by means of 'exploitation'.

Reinforcement learning algorithms are typically divided into two main categories, on-policy and off-policy algorithms. On-policy algorithms, such as SARSA, apply a 'learning on the job' approach where the algorithm learns about the actual policy it is performing. Off-policy approaches, such as Q-learning, apply a 'looking over someone's shoulder' approach where the algorithm learns an optimal policy by performing a different behavior policy [41]. Both on-policy and off-policy methods typically apply temporal-difference (TD) learning. TD-learning applies temporal difference updates seen in equation 2.36 to estimate the value function.

$$v(s) \leftarrow v(s) + \alpha \left(R(s,a) + \lambda v(s') - v(s) \right) \tag{2.36}$$

where: α is the learning rate applied on the TD-error: $(R(s,a) + \lambda v(s') - v(s))$

However, in model-free approaches, it is unknown how states transition into the next states so the state-action pairs are required to evaluate the policy. In an on-policy approach such as SARSA, the state-action pairs get updated through TD-learning by implementing the current policy π. The next action a' is sampled from the current policy π as seen below:

$$q(s,a) \leftarrow q(s,a) + \alpha \left(R(s,a) + \lambda q(s',a') - q(s,a) \right) \tag{2.37}$$

A policy that is applied frequently in RL is epsilon-greedy. A 'greedy' policy is one which always takes the action which leads to the best outcome, which is the action corresponding to the highest action value $q(s,a)$. A greedy policy does not consider the 'exploration-exploitation' trade-off. The agent needs to explore the environment to find a good estimate of each action value in order to find the optimal policy. Epsilon-greedy acts optimally with respect to the highest action values, except for a small probability epsilon [41].

Off-policy approaches, such as Q-learning, differentiate from this approach by sampling from a surrogate policy μ. It applies TD-learning on the current policy π by learning from the experience sampled from μ. Q-learning directly learns the optimal policy since it acts greedily while learning from an ϵ-greedy behavior policy.

$$q(s,a) \leftarrow q(s,a) + \alpha \left(R(s,a) + \lambda \max_{a'} q(s',a') - q(s,a) \right) \tag{2.38}$$

For both the on-policy and off-policy approaches, the same problem arises. The problem is calculating and saving each state-action value pair in tabular form. In complex environments, as are common for civil engineering structures, with large state spaces this problem quickly becomes unsolvable which has led to advances to approximate the action values by means of neural networks in a method called Deep Reinforcement Learning (DRL). It is also possible to directly approximate the policy with policy-based methods. The application of deep neural networks to approximate the action values or the policy will be the topic of the next subsection.

2.3.3. Deep Reinforcement learning (DRL)

The approach taken by reinforcement learning does come with its limitations. In reinforcement learning each state-action pair has its own value within a table, however, this is limited to small state and action spaces due to the curse of dimensionality. The problem is not only in saving all the pairs and tracking the table in its memory, it also requires data and time to be filled in accurately. The problem becomes one of generalization [41]. Not all states are visited as frequently, or not at all. The idea behind generalization is to extend the knowledge gained from a subset of states to the entire state space.

Generalization is applied in the form of function approximators. Examples of function approximators are linear approximators, decision trees, or neural networks. In DRL, deep neural networks (DNN) (deep referring to multiple layers) are used to approximate the function and will be the focus of this section. Neural networks consist of an input layer, one or more hidden layers, and an output layer. The layers are connected through weights and biases. The functions to be approximated in DRL are the

action-value function $q(s, a)$ or the policy function $\pi(s)$. An example is given below for the action-value function (equation 2.39).

$$q(s, a) \approx q(s, a|\Theta) \tag{2.39}$$

where: $q(s, a|\Theta)$ is the function computed by the deep neural network using Θ as the weights and biases in the neural network.

The cardinality of Θ is much less than all the possible state and action pairs, $|\Theta| << |S \cdot A|$. Also, changing a parameter $\theta \in \Theta$ consequently changes the value of the function for many other states. Therefore, information gathered by visiting one state can be generalized to many other states.

Many implementations of DRL algorithms exist, with the main distinction being value-based algorithms, policy-based algorithms, and actor-critic algorithms. In a broad way, the value-based algorithms apply a DNN to approximate the action-value function, the policy-based algorithms apply a DNN to approximate the policy function and actor-critic algorithms combine the advantages of both methods. The literature review performed in this subsection will be of a practical nature. The main types of algorithms will be discussed with their limitations and advantages. The subsection will be concluded by summarizing the current state of deep reinforcement learning within the inspection and maintenance planning of civil engineering structures. For more in-depth information about DRL, the overview written by Li is recommended [46].

Value-based methods

Value-based methods focus on approximating the value function of a certain state. Since RL adopts a model-free approach, the exact transition model of the environment is unknown making it difficult to work with the value function. For this reason, the action-value function is used which comes with the initial action and follows policy π afterward. By knowing all the action values for a given policy, the best policy is to act greedily with respect to its action values. An example of a value-based method is given in figure 2.5(a).

Another common value-based approach is using an advantage function $A_\pi(s, a) = q_\pi - v_\pi$. The advantage function expresses the relative advantage between actions and is similar to the baseline method of variance reduction applied in policy-gradient methods [47].

Deep Q-learning is the first and simplest application of using artificial neural networks as function approximators for reinforcement learning. DQN has been used by the team of DeepMind to outperform human professionals on 49 classic Atari games using only raw pixels as input [48].

DQN was able to bypass instability issues common when using off-policy methods with function approximators and bootstrapping by using experience replay and target networks [46]. The experience replay stores previous observation sequences (s, a, r, s') in its memory and is sampled uniformly to remove correlations in the data. The parameters from the target network are kept fixed and only updated periodically to reduce correlations between the q-values $q(s, a)$ and the target $r(s, a) + \gamma \max_{a'} q(s', a')$. Updating of the network parameters θ at iteration i is performed using the gradient of the following loss function $L(\theta)$:

$$L(\theta) = \left(r + \gamma \max_{a'} Q(s', a'|\theta_i^-) - Q(s, a|\theta_i) \right)^2 \tag{2.40}$$

where: $L(\theta)$ is the loss function to which gradient descent is applied to update the network parameters θ, θ are the parameters of the Q-network, θ^- are the parameters of the target network.

The work from Wei et al. shows an application of DQN to the structural maintenance of a bridge. A deep neural network can be used for many bridge cases with little changes to the architecture of the network [37]. In this work, the inspection and maintenance policy found by DQN is able to outperform both the time-based and condition-based benchmarks.

Since the introduction of DQN, many extensions to the algorithm have been made. One important improvement is the so-called Double-Deep Q networks (DDQN) which tackle the problem of DQN being prone to overestimation of the value estimates. Double-DQN (DDQN) uses both the target network and the Q-network to compute the target and has been found to successfully reduce the over-optimization [49].

Policy based methods

Policy-based methods directly seek to optimize the policy and do not maintain a model for the value function. An example of a policy-based method can be seen in figure 2.5(b). A parametrized formulation, e.g. a neural network, of the policy $\pi(a|s,\theta)$ is chosen, π_Θ, and is updated through maximizing the objective function, usually the expected return. Updating of the policy relies on policy gradient methods introduced by Sutton et al [50], where the policy parameters are updated proportional to the gradient ∇_θ and scaled with a learning rate α. The updating of the gradient is performed using gradient ascent. The policy gradient is given below:

$$\nabla_\theta \pi(a|s,\theta) = \mathbb{E}_{\pi_\theta}[_\theta log \pi(a|s,\theta) q_\pi(s,a)] \tag{2.41}$$

A Deep Policy Gradient method uses an artificial neural network to approximate the policy π. The input to the network is the current state and the output is a probability mass function over all possible actions.

In order to reduce variance in the gradients typically a baseline is subtracted from the action-value function, such as the value function $v(s)$ to form the advantage function $A(s,a)$ as explained under the value-based methods. The advantage function is used to express how favorable an action is compared to the mean value of that state. The algorithm REINFORCE applies the return r as an unbiased sample of the action-value function [51] using Monte-Carlo estimates, but learns much slower than RL methods with value functions [50]. Another approach is the use of actor-critic algorithms which apply a critic network to estimate the action values. The use of a critic network to estimate the value function reduces variance but can suffer from an increased bias by including additional approximations and assumptions in the learning process.

Policy gradient methods are able to handle high-dimension or continuous action spaces and can learn stochastic policies. Policy gradient methods usually also have better convergence properties than value-based methods. The downside of policy-gradient methods is that they tend to converge towards the local optimum and encounter high variance [46].

Actor-critic methods

Actor-critic methods lie at the intersection between both policy-based and value-based methods. The actor-critic consists of the 'actor' who tries to optimize the policy on the feedback it gets from the 'critic'. The actor updates the parameters of the actor-network in the direction suggested by the critic. An example of an actor-critic method can be seen in figure 2.5(c).

One of the first successful implementations of an actor-critic network using deep neural networks was asynchronous advantage actor-critic (A3C) by Mnih et al, [52]. The gradient of the objective function $\nabla\pi(a|s,\theta)$ is given below:

$$\nabla \pi(a|s,\theta) = \mathbb{E}_{\pi_\theta}[\nabla_\theta log_\pi(a|s,\theta) A(s,a,\theta,\theta_v)] \tag{2.42}$$

where: $A(s,a,\theta,\theta_v)$ is an estimate of the advantage function, θ are the parameters of the actor-network, θ_v are the parameters of the critic-network and $\nabla_\theta log_\pi(a|s,\theta)$ is known as the score function.

The estimate of the advantage function applied in A3C combines both samplings of the return and approximating of the value function using a critic network to make a trade-off between bias and variance. The advantage function is given below:

$$A(s,a|\theta,\theta_v) = \sum_{i=0}^{k-1} \gamma^i r + \gamma^k v(s|\theta_v) - v(s|\theta_v) \tag{2.43}$$

where: k is the length of the sampled return and $v(s|\theta_v)$ is the estimate of the value function of the critic network

Updating of the network is performed once the environment reaches a terminal state or at the end of the epoch t_{max}. The asynchronous methods were found to be learning significantly faster than DQN and outperformed all three value-based methods [52]. A3C, however, has poor data efficiency and

lacks robustness.

Trust region methods have been designed to stabilize policy optimization by introducing constraints to the gradient updates. The first implementation of a trust region was in the work of Schulman et al [53], which introduced the algorithm Trust Region Policy Optimization (TRPO). In TRPO, a surrogate objective function (see equation 2.44) is maximized under a constraint on the size of the policy update (see equation 2.45).

$$\max_{\theta} \mathbb{E}\left[\frac{\pi_\theta(a|s)}{\pi_{\theta_{old}}(a|s)} A(s,a)\right] \tag{2.44}$$

$$\text{subject to: } \mathbb{E}\left[KL[\pi_{\theta_{old}}(\cdot|s), \pi_\theta(\cdot|s)]\right] \leq \delta \tag{2.45}$$

where: θ_{old} refers to the policy parameters of the neural network before the update, KL is the Kulllback-Leibler divergence and δ is the trust region parameter.

By implementing trust regions, TRPO has been able to attain reliable performance while maintaining data efficiency. Although, TRPO is also relatively complicated and unable to handle algorithms with noise or parameter sharing. An updated version of TRPO is Proximal Policy Optimization (PPO) [54], which implements a new objective of clipped ratios and alternates between data sampling of the policy and policy optimization. PPO has been able to maintain the stability and reliability of trust-region methods such as TRPO and at the same time make the implementation much simpler and achieve better performance. In equation 2.47, the objective function of PPO is given.

$$r_t(\theta) = \frac{\pi_\theta(a|s)}{\pi_{\theta_{old}}(a|s)} \tag{2.46}$$

$$L^{CLIP}(\theta) = \mathbb{E}\left[\min(r_t(\theta)A(s,a), clip(r_t(\theta), 1-\epsilon, 1+\epsilon)A(s,a)\right] \tag{2.47}$$

where: $L^{CLIP}(\theta)$ is the main objective function of the algorithm, r_t is the probability ratio between the old and new policy, ϵ is a hyper-parameter to constrain big updates to the policy

Extensions for structural maintenance

With the recent successes of DRL algorithms described in this section in optimal control, specific applications to the structural maintenance domain have also been explored. The first application to be discussed is Deep Centralized Multi-agent Actor-Critic (DCMAC) proposed by Andriotis and Papakonstantinou [8]. The algorithm has been designed specifically for the optimization of stochastic control of multi-component engineering systems with large state and action spaces. DCMAC is an off-policy actor-critic algorithm that employs experience replay. The algorithm employs a centralized value function for the entire system (critic network) and multiple agents represented by a centralized actor-network with a factorized output for the different components. DCMAC makes it possible to include multiple component systems due to conditional independence between component actions, which makes the actor's output scale linearly with the number of components and effectively tackles the curse of dimensionality.

The success of DCMAC in incorporating linear scaling for multiple components is mainly due to the decentralization of the policy outputs. Implementation of decentralizing alleviates the curse of dimensionality and reduces the amount of computational complexity. Andriotis and Papakonstantinou continued this trend by introducing Deep Decentralized Multi-agent Actor-Critic (DDMAC) [45]. In DCMAC each control unit had full knowledge of the entire state-space of the environment and applies the same actor network for each control unit. In DDMAC the decentralization also applies to the actor-network, effectively having an actor-network for each control unit with no parameter sharing. Therefore each control unit acts optimally with respect to its own states but is judged according to the value of the entire system.

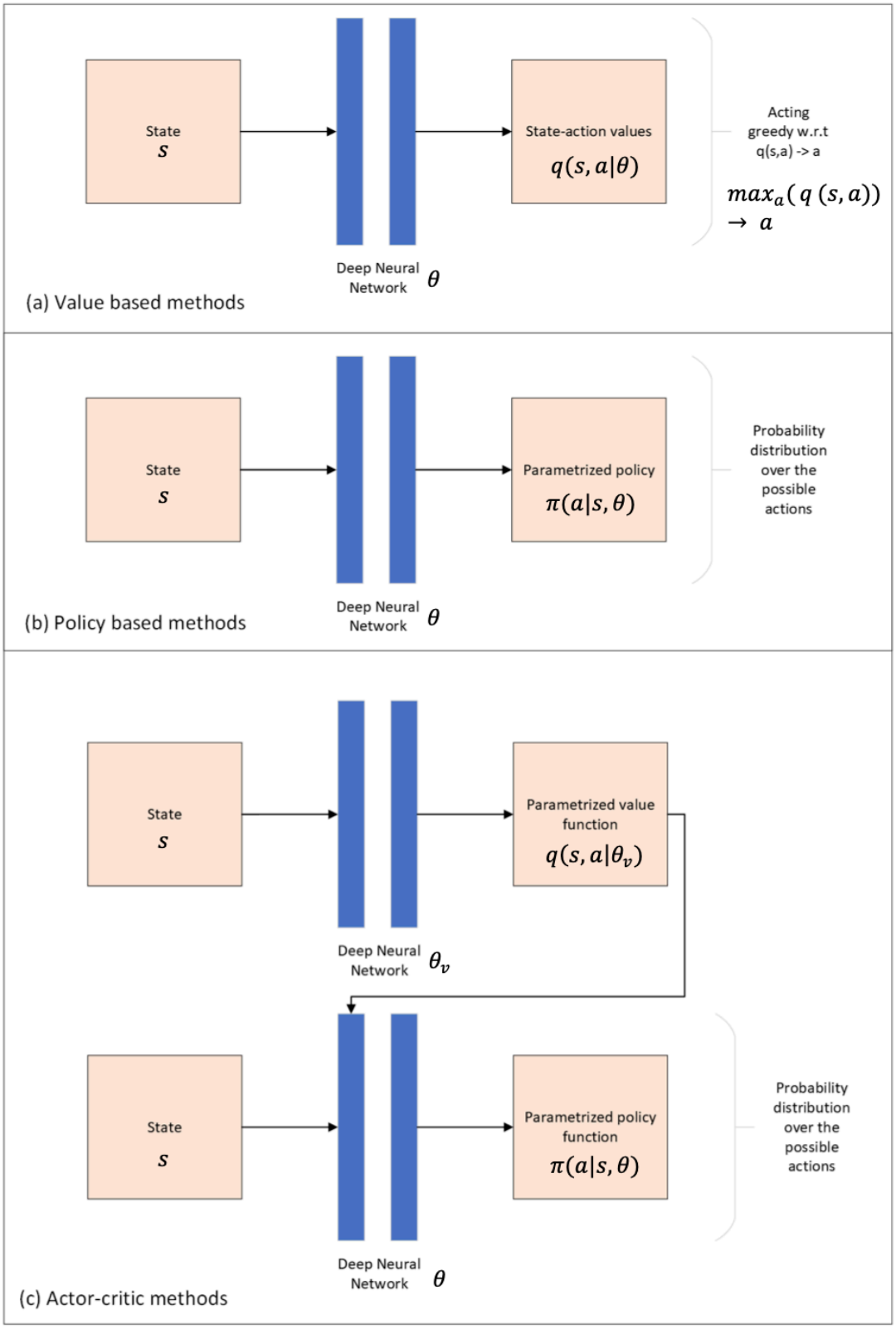


Figure 2.5: Main categories of Deep Reinforcement learning algorithms (a) Value-based methods, (b) Policy-based methods, (c) Actor-Critic methods

2.4. Current state of research

The literature review touched on three main topics which fall under the umbrella of ”planning of structural maintenance under a changing climate”. The goal of the literature review was to have a broad foundation on the three main categories to create a workflow for an adaptive inspection and maintenance framework.

From the existing literature, it was found that many aspects of climate change lead to reductions in safety and increases in costs, varying from more extreme natural events to reductions in durability. Changes in climatic conditions have a direct link with decreases in durability which can lead to unsafe situations. Especially the effects of different climate scenarios on corrosion have been discussed in many studies due to a direct reduction in safety due to degradation and the high global costs of corrosion.

Since a link has been established between climate change and the increased deterioration of reinforced concrete structures, the question arises of how to adapt to such situations. Due to the many effects of climate change on our infrastructures, inspection and maintenance planning is required, which can adapt to changes and make custom policies. Wang et al. [6], touch on this subject by stating that existing structures for which design did not take into account climate change will likely suffer from decreased durability. They recommend finding effective adaptation options for individual structures and increasing monitoring and maintenance.

A new type of inspection and maintenance planning in the form of Deep Reinforcement Learning has been explored which is able to operate in a model-free environment with imperfect information about its environment. Recent studies have provided algorithms that are scaleable to real-world engineering problems [8], [39].

2.5. Bridging the gap

This study focuses on bridging the gap between two theoretical topics: climate change and deep reinforcement learning. Traditional heuristic-based approaches require perfect information about the environment which is uncommon in climate change. DRL-based I&M policies are a way to deal with stochastic uncertainty and find an optimal policy by implementing a belief in the state of the environment. The bridge to be explored between both topics is provided by developing a framework for dynamic inspection and maintenance planning which updates under imperfect information on the current state of climate change, the condition of a civil engineering structure, and inspection/maintenance action outcomes.

3

Strategy of the framework

The literature review shows that our infrastructure is deteriorating at an accelerated rate due to climate change in a multitude of risk categories. Due to changes in the deteriorating environment, our infrastructure is prone to reductions in safety and an increase in life-cycle costs. The use of DRL algorithms in the sequential decision-making of inspection and maintenance plans has a big potential for advantages in terms of life-cycle costs and safety under a model-free environment with uncertain action and observation outcomes.

This section explores the incorporation of a probability distribution over different climate scenarios as an input for sequential decision-making. An optimal policy is sought for using a DRL approach due to the large state and action spaces which are common for complex civil engineering structures. In figure 3.1 the problem is dissected in components. Underneath the figure, a brief description of each component is provided. Inspiration for the dissection of the framework into its distinct components has been gathered from [55].

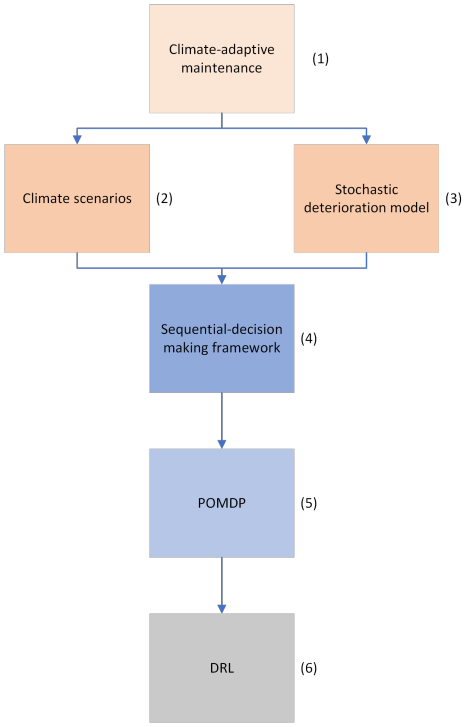


Figure 3.1: Components of the framework

1. The strategy of the framework is to have an inspection and maintenance policy that is adaptive to changes in the climate by incorporating a belief in the possible climate scenarios.
2. The climate scenarios can be used to estimate how a climate parameter will change in the future. An example of a climate parameter can be temperature and relative humidity on the degradation side, and sea-level rise or extreme weather events on the loading side. The real scenario is somewhere in between the different climate scenarios. The climate scenarios provide a method to plan for different future scenarios. A method to implement the different scenarios is the implementation of a belief in the climate scenarios. The belief is a probability distribution over the possible climate scenarios as input for the sequential decision-making framework.
3. The resistance of the structure is decreasing over time due to deterioration. The deterioration calls for maintenance of the infrastructure, however, it is difficult to plan the maintenance actions due to the many uncertainties involved in the stochastic deterioration process.
4. Finding the optimal sequence of inspection and maintenance actions of a deteriorating infrastructure is the main problem. This component introduces the novelty of the work where both climate change and deterioration are considered within finding an optimal sequence of actions under sequential decision-making. The incorporation of different scenarios of a climate parameter in the sequential decision-making framework can be seen in figure 3.3.
5. Partially observable Markov decision processes are used as a mathematical framework to formulate sequential decision-making when actions and observations have uncertain outcomes [40]. A brief explanation of the components of a POMDP is given below in figure 3.2.
6. Deep reinforcement learning algorithms have been shown to efficiently solve POMDPs as has been shown in section 2.3.3. The result of a DRL algorithm is an optimal and adaptive inspection and maintenance policy with respect to the expected life-cycle costs of the structure.

A graphical representation of the underlying POMDP for three time steps is given below in figure 3.2. The figure describes the interaction between the belief **b**, hidden states s, observations o, actions a, and rewards r.

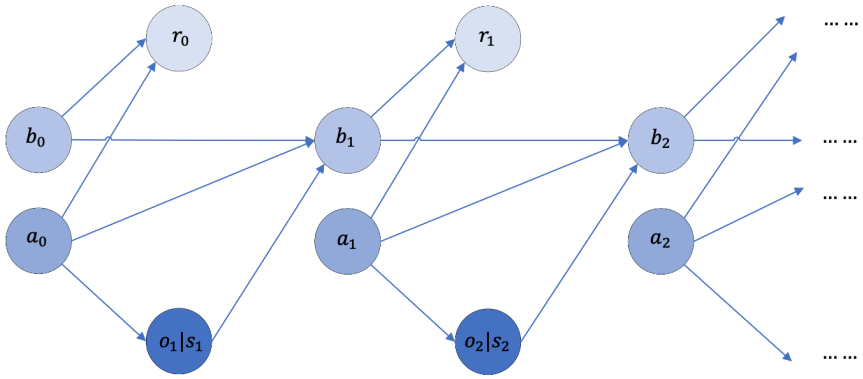


Figure 3.2: Graphical representation of the POMDP framework

- **b** is the probability distribution over the hidden state $s_t \in S$ of the environment. In this framework, the belief has a probability distribution over the state of the environment (e.g. damage of a component) and the state of climate change (the possible climate scenarios). Updating the belief over the climate scenarios happens at each decision step and an update over the belief of the environment occurs if an action is performed. The state space is discretized to significantly decrease the computational time.
- $o_t|s_t, a_t \in O$ is the observation performed at each decision step, which depends on the hidden-state s_t. The observation includes an update of the climate parameters and an update on the state of the environment if an action is performed. Updating of the belief is done according to Bayesian updating.

- $a_t \in A$ is a vector of actions for each component taken at the decision step. The actions consist of an inspection and/or maintenance action per decision step.
- $r_t \in R$ is the reward given by the environment to the agent for taking action a_t whilst in belief-state b_t. In the field of structural maintenance, typically costs are used as 'negative rewards'.

The incorporation of the climate scenarios in the agent-environment interaction can be seen in figure 3.3.

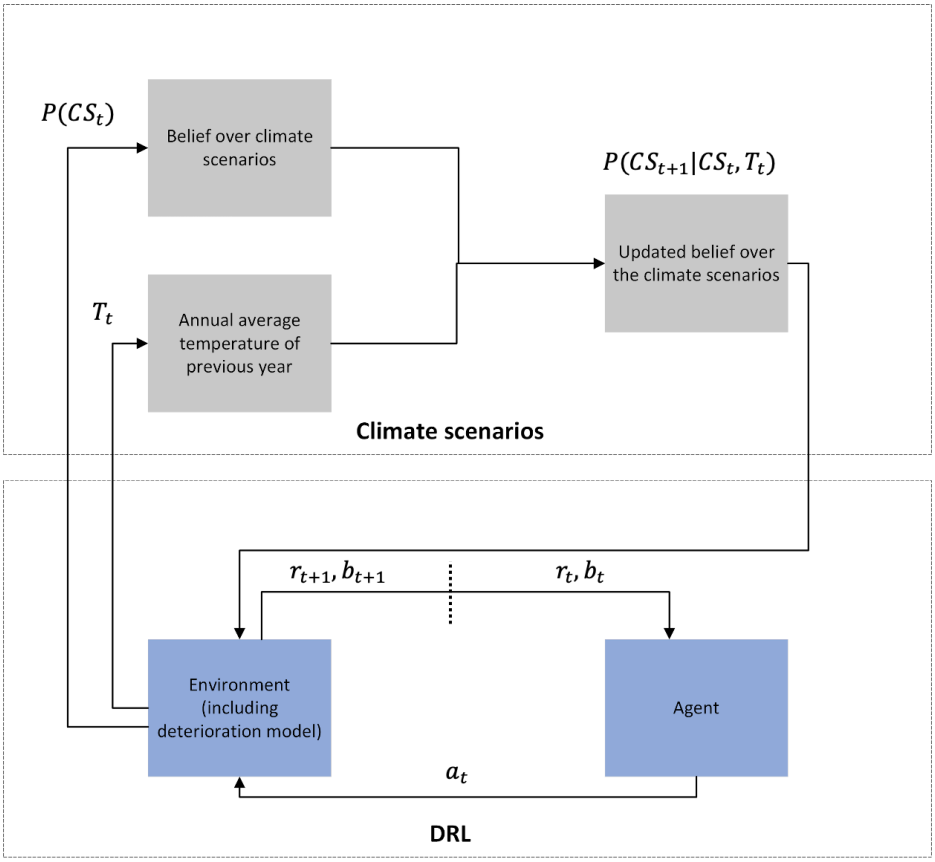


Figure 3.3: Incorporation of climate scenarios in the agent-environment interaction.

where: $P(CS_t)$ is the probability distribution over the different climate scenarios at time step t, T_t is the temperature at time step t, $P(CS_{t+1}|CS_t, T_t)$ is the updated probability distribution over the climate scenarios for time step $t+1$.

The layout of the strategy of the framework is inspired by the work performed by Rupke [56]. In this work, the framework for risk-informed maintenance is split into three steps: a preparation step, a processing step, and an optimization step. In this thesis, the steps are translated toward the setup of the physical model, the setup of the POMDP, and the finding of the optimal policy. The actions required per step are changed since the problem itself and the solution method are different.

The first step, the preparation, is the setup of the physical system. It is used to set the goal of the framework, identification of the deterioration mechanisms, collect the required data, incorporation of scenarios of the governing climate parameter, and description of inspection and maintenance actions. The second step, processing, translate the physical system to a POMDP to formulate the sequential decision-making framework and includes the setup of the states, actions, rewards, a transition model, and an observation model. The third and final step is the application of a DRL algorithm to solve the POMDP and find the optimal inspection and maintenance sequence for the infrastructure under uncertain observation and action outcomes under a changing climate.

3.1. Setup of the physical system

In this section, the physical system is set up by explaining the following four sub-steps. First, the goal of the inspection and maintenance plan is determined and requirements are set up. Next, the physical system is described with a stochastic deterioration process and scenarios of the governing climate parameter are set up. Followed by the input data for the deterioration mechanisms, climate model, and infrastructure. The setup is concluded by determining the inspection and maintenance actions to be considered in the planning of the optimal inspection and maintenance sequence. A flowchart of the first step is given in figure 3.4.

3.1.1. Goals and requirements

Requirements are to be determined to formulate a clear goal for the inspection and maintenance policy. Requirements are set up to frame the policy to the problem. In this framework, the requirements described in [56] are used which consist of functional, cost, and time requirements.

Functional requirements describes the services to which the optimization must comply. This consists of a combination of safety (ULS) and serviceability (SLS) demands.

- *Probability of failure, ULS:* The inspection and maintenance policy must prevent failure of the structure to occur. This is typically assessed on the probability of failure of the structure. This can be in the form of implementing a boundary for a certain probability of failure per decision step or penalizing the policy for higher probabilities of failure.
- *Serviceability, SLS:* The structure must be safe to use under actual service loads. Therefore, risks must be penalized which limits the serviceability to prevent the unavailability of the structure. Examples can be spalling of the concrete or big deflections of the structure.

Economical requirements The costs of the bridge are described by the inspections, maintenance actions, failure probability, availability, and serviceability of the bridge. The combination of the costs describes the life cycle costs of the bridge and is used to find the optimal sequence of inspection and maintenance actions.

Time requirements are used to bind the policy towards a certain time frame for the policy. The time frame can be seen as the remaining service life of the structure.

The requirements listed above have to be formulated as a goal for the inspection and maintenance plan. The goal is used to set up and verify the strategy of the framework. An example of such a goal is given below.

Find an inspection and maintenance strategy considering climate change which holds

- that the probability of failure remains below 1% per year and the associated risks are minimized
- and the life-cycle costs are lower than the heuristic-based benchmarks
- within the time frame from 2000 to 2100 ($t_{life} = 100 years$).

3.1.2. The physical model

The physical model consists of identifying the critical components of the structure, a degradation model for the critical components, the effect of different climate scenarios on the resistance and loading of the environment, and a definition of the limit states.

Identification of the critical components is done by performing a risk analysis. The risk analysis follows what parts of the structure are governing the failure probability and how they constitute to the failure probability. The components which follow from the risk analysis are used to design an inspection and maintenance policy. An example of a risk analysis is Failure Mode Effect and Criticality Analysis (FMECA) [56].

Once the critical components are identified, it is important to find the governing deterioration mechanisms of the components. The deterioration makes the resistance time-dependent and is necessary to plan accordingly. The effect of climate scenarios can be incorporated on both sides of the limit state. For the deterioration of the environment by for instance accelerated corrosion, or, for the loading, in

terms of frequency and intensity of extreme weather events. An example of the deterioration of reinforced concrete is covered in the literature review. In the literature review, the deterioration mechanism can be seen in section 2.2.2 and the climate scenario is described in section 2.1.1.

Next, both the governing deterioration mechanism and the climate scenarios are used to set up a stochastic deterioration process for each climate scenario. The result is a limit state which depends on the belief over the possible climate scenarios in terms of loading and resistance.

The limit state function Z compares the resistance R with the loading S to assess whether failure will occur (see equation 3.1). If the limit state function is positive ($Z > 0$), no failure will occur. A negative limit state function ($Z < 0$) indicates failure. A zero limit state function ($Z = 0$) indicates the limit state boundary. Assessment of the limit state can be done using level I, II, and III methods such as First Order Reliability Method (FORM, level II) and Monte Carlo analysis (level III). In this thesis, only Monte Carlo will be used which approximates the failure probability by sampling from stochastic distributions and counting the times that failure occurs.

$$Z = R - S \tag{3.1}$$

The critical components of the physical structure are identified, the limit state of the structure is time and climate dependent and a method exists to assess the probability of failure for the structure. The physical model is therefore identified sufficiently to transition towards the next step, the input data.

3.1.3. Input data

The formulation of the limit state shows the required input parameters. The input parameters consist of stochastic distributions of the geometry, materials, deterioration mechanisms, loading, and climate. Values can be obtained either through experiments or existing literature. In this research, only stochastic distributions from existing literature will be used.

The result of this step is a list of the input variables, corresponding stochastic distribution, and required parameters for the distribution. The input data also includes the scenarios of the governing climate parameters for the given time frame listed in the goal.

3.1.4. Inspections and maintenance actions

A goal for the inspection and maintenance strategy is set, the structure is broken down into critical components, and a limit state function is set up to link the resistance of the critical components to the loading to find a probability of failure for the given physical model. The next step is to set up the possible inspection and maintenance actions to find the optimal sequence of actions. The end of this phase is a list of all possible inspection and maintenance actions to be performed with their corresponding description, cost, and consequence for the state of the infrastructure. The actions are indexed from least expensive up to most expensive in the following format: 'Maintenance action 1 = m_1', 'Maintenance action 2 = m_2', 'Inspection action 1 = i_1', 'Inspection action 2 = i_2' etc. An illustrative example of such a list for a structure subjected to chloride-induced corrosion can be seen below in table 3.1.

Action	Description	Costs [€]	Consequence
Minor repair, m_1	The cracked concrete is restored and rebar is cleaned of corrosion	1000	The deterioration rate of the corrosion is reset ($\tau_t = \tau_0$)
Replacement, m_2	The cracked concrete is restored and the rebar is replaced	5000	The deterioration rate and damage are reset ($\tau_t = \tau_0, d_t = d_0$)
Inspection 1, i_1	A visual inspection is performed and the structure is rated on a scale of 1-10	100	Damage state is found with probability $p = 0.60$
Inspection 2, i_2	An inspection is performed measuring the corrosion current density of the rebar.	500	Damage state is found with probability $p = 0.90$

Table 3.1: Example list of I&M actions

3.2. Setup of the POMDP

After the preparation of the physical model is finished, the physical model is translated towards a POMDP as a framework for sequential decision-making under uncertain outcomes of observations and actions. The setup of the POMDP is finalized by translating the physical model towards the components of the mathematical framework of the sequential decision-making problem, the POMDP. The components $< S, A, R, O, Z, T, \gamma >$ include states $s_t \in S$, actions $a_t \in A$, rewards $r_t \in R$, observations $o_t \in O$, an observation model $Z = P(o_{t+1}|s_{t+1}, a_t)$, a transition model $T = P(s_{t+1}|s_t, a_t)$ and a discount factor γ.

3.2.1. Introduction of POMDP framework

In this step of the processing phase, sequential decision-making is formulated as a POMDP. This includes a physical definition of each component, a graphical model of the POMDP, and a clear definition of the state space. The introduction should formulate the POMDP so it is clear how each component is related to the other. After this step, it should be clear what components need to be determined and how the environment transitions according to a decision step.

The POMDP is accompanied by a belief-state **b** which is a probability distribution of the states S. An example of the belief-state for a system described by a deterioration rate τ and damage state s per component k and possible climate scenarios CS for the system is given below in equation 3.2.

$$\mathbf{b} = \begin{cases} [s_0, s_1, ..., s_n]_k & n \text{ is the amount of discretized states}, k \text{ is number of components} \\ [\tau_0, \tau_1, ..., \tau_k] & k \text{ is number of components} \\ [CS_0, CS_1, ..., CS_j] & j \text{ is the number of the climate scenarios} \end{cases} \tag{3.2}$$

where: b is the belief state, which is the probability distribution over the hidden states shown to the right of the curly bracket.

3.2.2. Degradation

The deterioration is described by a stochastic deterioration process, such as a gamma process as described in section 2.2.3. The result of the gamma process is random positive damage increments for each component for each decision step. If the resistance depends on the possible climate scenarios then a stochastic deterioration process has to be set up for each. The input parameters for the gamma process follow from a random value degradation model such as the one presented in section 2.2.2.

The stochastic deterioration process is used to set up the transition model of the environment $T \leftarrow P(s'|s, a)$. The underlying continuous damage state of the environment is discretized to significantly decrease the computational time required. The transition model is acquired by drawing numerous random samples of the gamma process for a climate scenario and calculating the number of times that state transitions occur. An example of a transition model can be seen in equation 3.3 for a state-space

discretized in five states.

The transition model describes the update of the belief state after a decision step. The probability of being in a certain state, or, the belief over the state of the environment, is used to calculate the failure probability of the structure in the next step.

$$P(s_{t+1}|s_t) = \begin{bmatrix} p(s_0|s_0) & p(s_1|s_0) & p(s_2|s_0) & p(s_3|s_0) & p(s_4|s_0) \\ 0 & p(s_1|s_1) & p(s_2|s_1) & p(s_3|s_1) & p(s_4|s_1) \\ 0 & 0 & p(s_2|s_2) & p(s_3|s_2) & p(s_4|s_2) \\ 0 & 0 & 0 & p(s_3|s_3) & p(s_4|s_3) \\ 0 & 0 & 0 & 0 & p(s_4|s_4) \end{bmatrix} = \begin{bmatrix} 0.90 & 0.05 & 0.03 & 0.02 & 0.00 \\ 0.00 & 0.90 & 0.05 & 0.03 & 0.02 \\ 0.00 & 0.00 & 0.90 & 0.05 & 0.05 \\ 0.00 & 0.00 & 0.00 & 0.90 & 0.10 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{bmatrix} \tag{3.3}$$

$$\sum_{i=0}^{n} P(s'|s_n) = 1 \tag{3.4}$$

where: $P(s_{t+1}|s_t)$ is the transition model for the deteriorating environment, $p(s_i|s_j)$ gives the probability of going to state s_i given that the system is in state s_j. The elements below the diagonal are zero since a component cannot transition toward an improved state for positive damage increments.

3.2.3. Failure probability

In the previous step, the transition model was determined. With the transition model, the transition of the state of the environment over time is known. The degradation of each component can be used to determine the failure probability of the structure. The degradation is described by the belief over the state of each component at each decision step t. The failure probability can be acquired by applying the limit state function to the damage state. By applying the law of total probability, conditional failure probabilities $P(F, i|s_j)$ (Probability of failure of component i, given that it is in state s_j) can be combined with the belief-state to form a failure probability $P(F)$ per component i:

$$P(F, i) = \sum_{j=0}^{n} P(F, i|s_j) \cdot P(s_j) \tag{3.5}$$

where: $P(F_i)$ is the probability of failure for component i, j is the index of the state, $P(F, i|s_j)$ is the probability of failure for component i given state j, $P(s_j)$ is the probability of being in state s_j which follows from the belief-state **b**.

The conditional failure probabilities $P(F_i|s_j)$ are found by applying a numerical reliability analysis such as the Monte Carlo analysis. Analytical methods are typically unable to find a solution due to the complexity of the limit state. The failure probability is computed at each time step using the conditional failure probability and the belief state. The failure probability per component can then be used to compute the failure probability of the structure according to the influence of each component which follows from the risk analysis. The failure probability of the structure is used in the formulation of the rewards by including risk costs or boundaries in the decision-making.

3.2.4. Actions and rewards

With the deterioration environment known and the failure probability determined, the next step is the description of the inspection and maintenance actions. The description includes the formulation of the transition model and observation models for each action and the setup of the reward function.

The actions consist of inspection actions and maintenance actions. Each action affects the environment in a different way. A maintenance action improves the condition of the structure and directly affects the hidden-state S. An inspection action identifies the hidden state of the environment according to the observation model Z and updates the belief state of the agent **b** according to an observation o.

Maintenance actions have to be formulated with a transition model $P(s'|s, a)$ which described how the state s is updated to a successor state s' after an action a is performed. The inspections have to be formulated with an observation model $P(o|s, a)$ which is used to update the belief state after a

decision step. An example of a transition model which replaces a component (resets the state s to the undamaged state s_0) is given below in equation 3.6.

$$P(s'|s, a_{replace}) = \begin{bmatrix} p(s_i|s_0) \\ p(s_i|s_1) \\ p(s_i|s_2) \\ p(s_i|s_3) \\ p(s_i|s_4) \end{bmatrix} \begin{bmatrix} 1.0 & 0 & 0 & 0 & 0 \\ 1.0 & 0 & 0 & 0 & 0 \\ 1.0 & 0 & 0 & 0 & 0 \\ 1.0 & 0 & 0 & 0 & 0 \\ 1.0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.6}$$

where: $P(s'|s, a_{replace})$ is the transition model if a replacement is performed, $P(s_i|s_0)$ is the probability of going to successor state s_i if in state s_0.

The observation model describes the probability of making observation o given the hidden-state s and action a. An example of an observation model which measures the hidden-state s_j with probability p and adjacent states with probability $(1-p)/2$ is given below:

$$O = \begin{bmatrix} p & 1-p & 0 & 0 & 0 \\ (1-p)/2 & p & (1-p)/2 & 0 & 0 \\ 0 & (1-p)/2 & p & (1-p)/2 & 0 \\ 0 & 0 & (1-p)/2 & p & (1-p)/2 \\ 0 & 0 & 0 & 1-p & p \end{bmatrix} \tag{3.7}$$

where: p is the level of precision of measuring the correct hidden state.

With the actions properly described and the failure probability known, the reward function can be formulated. As stated before, the reward function $r(s, a)$ depends on the state of the environment and actions chosen at each decision step. The rewards consist of maintenance costs C_{action} and risk costs C_{risk}. Risk is calculated based on the product of the cost of an event C_{event} and the probability of an event P_{event}. Action costs consist of the costs of performing an action per component $C_{action,i}$.

$$C_{risk} = P_{event} \cdot C_{event} \tag{3.8}$$

$$C_{action} = \sum_{i=0}^{n_{comps}} C_{action,i} \tag{3.9}$$

The life-cycle costs C_{tot} are used to compare policies and are described based on the discounted reward over the lifetime of the structure. Since the deterioration is a stochastic process, the expected value is used (see equation 3.10

$$C_{tot} = \mathbb{E}[G_t] = \sum_{i=0}^{T} \gamma^k (C_{risk} + C_{action}) \tag{3.10}$$

where: C_{tot} are the expected life-cycle costs, G_t is the return,γ is the discount factor and T is the length of the planning horizon.

3.3. Finding the optimal policy

This framework section covers the steps required to find an optimal policy. The life-cycle costs described in the previous section are translated toward an objective for the decision-maker. Heuristic-based benchmarks are set up for validation of the results and to compare the optimal policy with traditional approaches in terms of safety and costs. The optimal policy is found by applying a DRL algorithm to the POMDP described in the previous section. After applying the algorithm to the environment and training the model, results are found. The DRL algorithm should be described sufficiently for the results to be reproducible, this includes giving pseudo-code, hyper-parameters, and the training process. The framework is finalized by describing the found policy and comparing the policy with the benchmarks.

3.3.1. Objective function

The optimal policy π^* maps the actions to the states in such a manner that the lowest life-cycle costs are found. An objective function is required to optimize and compare the different policies. The formulation below is used to describe the objective of the decision maker, see equation 3.11.

$$\pi^* = \arg\min_{\pi \in \Pi} [C_{tot} | a \leftarrow \pi(\mathbf{b}_0, a), \mathbf{b}_0] = \arg\min_{\pi \in \Pi} v^{\pi}(\mathbf{b}_0) \tag{3.11}$$

where: $\mathbf{b}_0$ is the initial probability distribution over the states, v^{π} is the value function giving the expected discounted life-cycle costs following policy π.

3.3.2. Benchmarks

In theory, DRL algorithms converge toward an optimal policy. However, the real advantage of the optimal policy can only be shown if compared to a traditional approach. Benchmarks are required to evaluate the policy in terms of safety and costs.

Traditional inspection and maintenance planning is based on heuristic approaches. A Heuristic approach is typically used when finding the optimal policy is too complex and an approximation is considered sufficient. Optimization in heuristic-based approaches is based on predefined heuristic rules to limit the total amount of explored policies. The optimality of such an approach depends greatly on the designer's experience with such problems [39].

Based on a heuristic rule, many simulations need to be performed for the different combinations of intervals to find converged values of the life-cycle costs. Below is described how a benchmark could be defined. An example of how a converged benchmark should be described is given in table 3.2.

- Define a heuristic-based decision rule
- Define a grid of intervals for the decision rule for each action
- Perform simulations of the environment for the different combinations of intervals.
- Provide the final results of the benchmarks in tabular form.

Benchmark	**action 1**	**action 2**	**inspection 1**	**inspection 2**	**Expected LC costs**
Time-based	5 years	15 years	-	-	20000 euro
Condition-based	5% damage	15% damage	2 years	6 years	18000 euro

Table 3.2: Example table of heuristic benchmarks

3.4. Flowchart of the framework

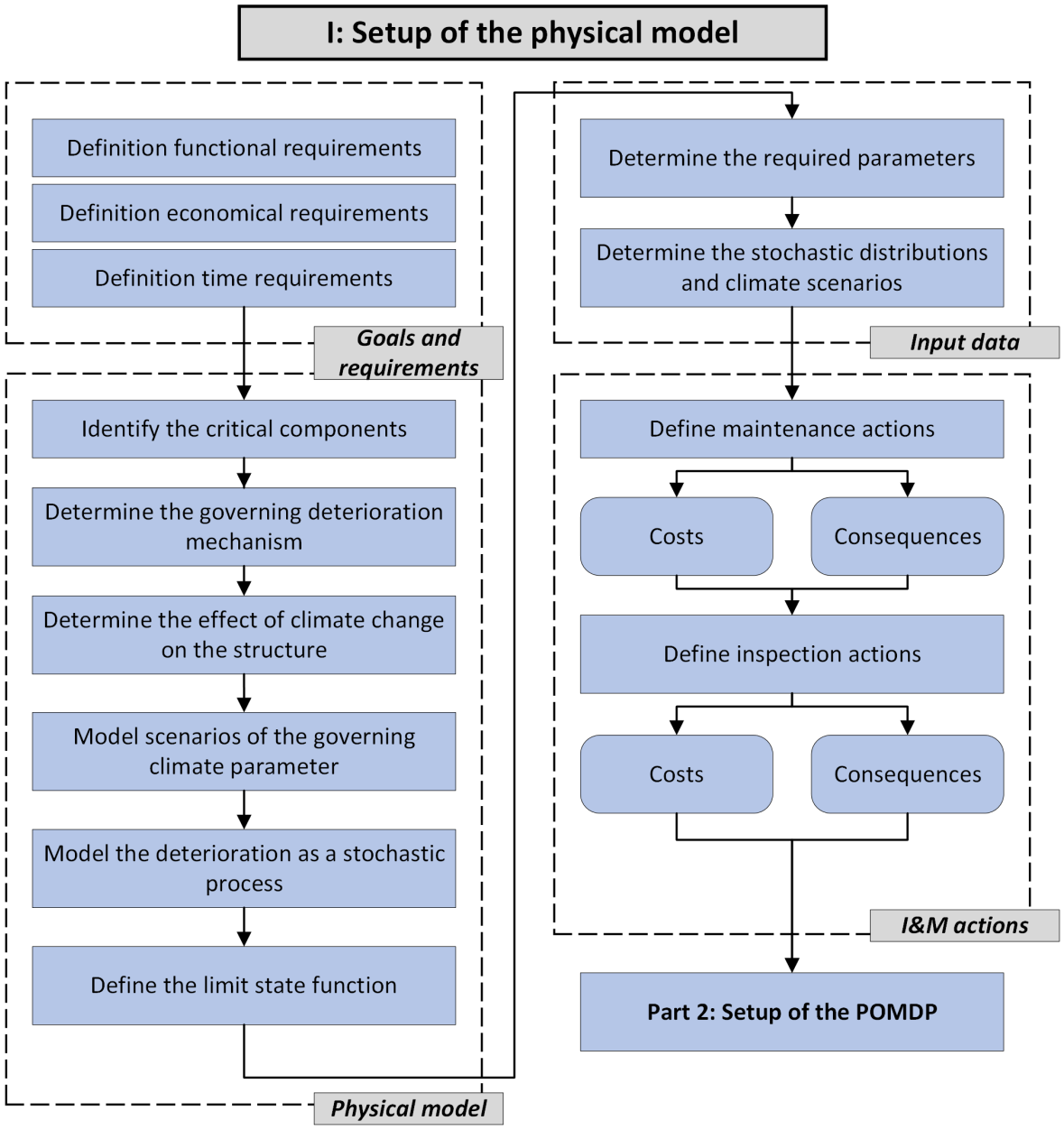


Figure 3.4: Flowchart of the physical model

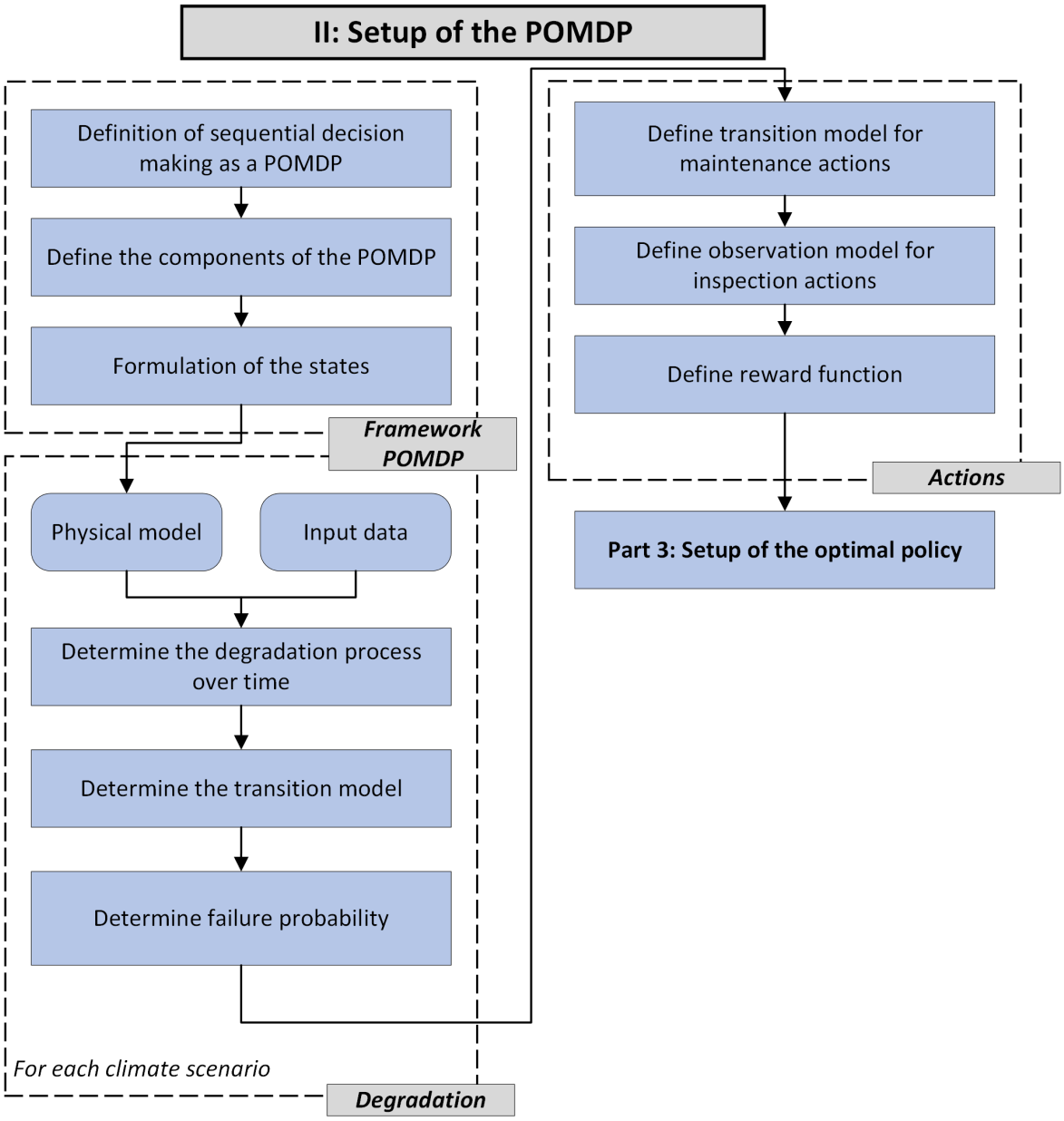


Figure 3.5: Flowchart of the POMDP

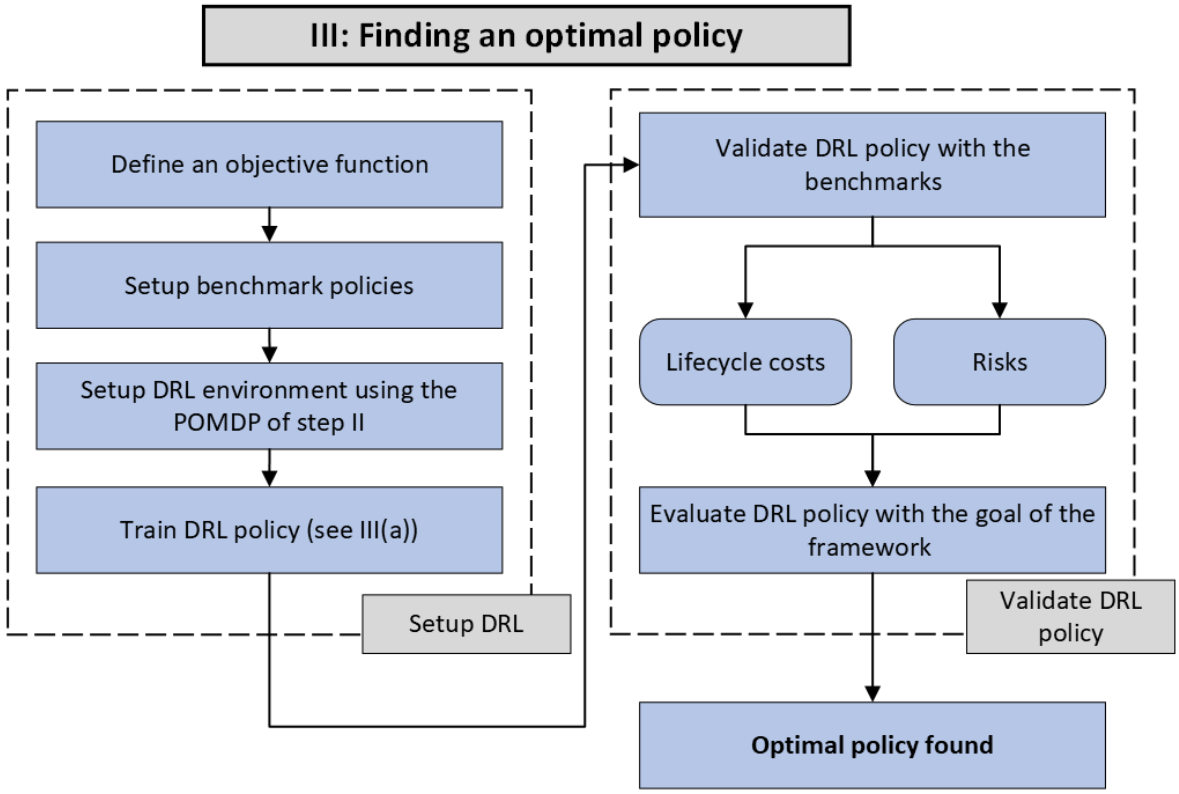


Figure 3.6: Flowchart of the optimal policy

Part II

Case study

4

Description

To show an application of the proposed framework, a case study will be presented as an illustrative example. The proposed case study is a three-span reinforced concrete slab bridge under maintenance by an illustrative client named Peter. Peter has performed a risk analysis on the illustrative case study and found the bridge to be subjected to chloride-induced corrosion which has initiated at the start of the inspection and maintenance plan. From the risk analysis, it follows that the longitudinal reinforcement of the bridge is governing the failure probability of the bridge. In this study, linear static analysis is used as a 'risk analysis' to find the hot spots of the highest bending stresses (see Appendix A. Assumptions stated by the 'client' have followed from the literature review in chapter 2. The assumptions are only a modeling choice to introduce the framework and can be changed according to the wishes of the user. The topics discussed in this chapter correspond with the first step of the framework, the setup of the physical model.

In the first section, a general description of the bridge will be given. In the general description, the critical components will be introduced and the deterioration mechanism is given. The governing climate parameter is introduced and the influence on the deterioration is described. The different inspection and maintenance actions are explained. The next section starts with some general modeling assumptions and is followed by assumptions taken for the deterioration process and the climate scenarios. The chapter is finalized with tables summarizing the input parameters for the next step, the setup of the POMDP. The setup of the POMDP and finding of the optimal policy will be discussed in chapter 5: Analysis.

4.1. General information

The bridge examined in this case study is in close correlation with a classical deterioration study performed by Stewart [23]. Most of the input parameters are kept the same, however, some design parameters have been changed to be in compliance with the Eurocodes. The height of the concrete slab and reinforcement ratios have been changed in order to be in accordance with the loading (Eurocode 1 part 2: Traffic loads on bridges [57]) and the reliability of the structure (Eurocode 0: Basis of structural design [58]). The bridge no longer corresponds with the Rattle Snake Creek Bridge in Ohio used in the study [23] and will be described as an original illustrative example bridge. The design calculations can be found in Appendix A.

The bridge is a three-span reinforced concrete slab bridge built in the year 1970. The bridge is simply supported on abutments at the ends and by piers for the middle supports. Each pier is supported by seven RC piles assumed to sufficiently support the superstructure. The width of the bridge is 11 m and consists of two opposing driving lanes. The two end spans are 9.83m and the middle span is 12.2m. The height of the concrete slab is 532 mm with a wearing course of 19mm, a top cover of 46mm, and a bottom cover of 32mm.

The client named Peter wishes to extend the service life of the bridge by 30 years by implementing a new inspection and maintenance plan. At the start of the inspection and maintenance plan, Peter

performed an (illustrative) Failure Mode and Effects, Criticality Analysis (FMECA) to identify the critical components. The risk analysis was accompanied by a concrete resistivity measurement and a half-cell potential measurement. The measurements have led to assumptions about the corrosion which will be stated in this section.

Over the years chlorides have penetrated the concrete and accumulated to a critical level for corrosion to initiate at the longitudinal reinforcement bars in the RC slab. Due to corrosion, the resistance of the structure is starting to decrease which can lead to cracking of the concrete and eventually leads to unsafe situations. The components found to be vital for the deterioration of the concrete slab are the longitudinal reinforcement bars at the critical cross-section of the highest bending moments. Five critical cross-sections have been identified to be the governing components for the risks described above. The critical cross-sections correspond to (i) the middle of the span (CS1, CS3, and CS5) for the positive bending moments and (ii) The negative bending moments above the supports (CS2, CS4).

The reinforcement detailing consists of a standard reinforcement net of $\varnothing 10 - 150mm$ in the lateral and longitudinal direction, top and bottom. In the middle of the first and last span, $\varnothing 20 - 140mm$ is added to the bottom reinforcement to support the peak positive bending moments. In the middle of the second span, $\varnothing 20 - 160mm$ is added to the bottom reinforcement to support the peak positive bending moments. At the middle supports, $\varnothing 20 - 105mm$ is added to the top reinforcement to support the peak negative bending moments. The standard reinforcement net is closed at the borders of the slab by applying hairpins to provide sufficient shear resistance and to close the concrete section to prevent cracking. The reinforcement detailing can be found in Appendix B.

Peter has many bridges in his care and found something interesting and costly to occur. Many of the bridges have been shown to deteriorate at an accelerated rate over the last decades and he speculates that changes in the climate are the cause. The client is an advocate for climate adaptive structures and wishes to implement a changing climate in a risk-informed inspection and maintenance strategy for the five components given above. According to the risk analysis performed by the client, the temperature seems to be the governing parameter for the changes in the deteriorating environment. From the measurement taken of the concrete slab, Peter has found that no corrosion has yet occurred at any component but that the probability of corrosion to initiate is high. This leads to the assumption that no corrosion damage has yet occurred but starts to propagate for each component at $t = 0$.

The safety of the structure is in danger due to a decrease in the effective cross-sectional area due to chloride-induced corrosion propagation. The serviceability of the structure is in danger since the corrosion causes expansive corrosion products to be produced which causes the concrete to crack and eventually spall. This leads to pieces of concrete breaking off and leads to unsafe situations and corrosion occurring at an even faster rate. The former risk is known as an ultimate limit state (ULS) and the latter as a serviceability limit state (SLS). Both limit states are assessed for each component i at each decision step t.

The concrete slab is seen as a series system. This means that if a component fails, the entire structure fails. In this thesis, it is assumed that no dependence exists between the components when failure is assessed. This is a simplified scenario and typically there does exist dependence between the different components, however, it makes the assessment of failure more complex and does not add anything to the conclusions of the research.

The inspection and maintenance strategy consists of the following actions. A non-destructive inspection action that measures the effective cross-section of the reinforcement for a component with a certain probability p. The probability p reflects the accuracy of the inspection method [43]. The maintenance actions consist of a minor repair action and a major repair action. The minor repair action replaces the concrete cover and cleans the corrosion of the reinforcement. This action resets the probability of concrete spalling P_s to occur and resets the deterioration rate τ, the damage of the reinforcement d however remains the same. The major repair action replaces the concrete cover and the deteriorated reinforcement is repaired, effectively resetting the corrosion process. The chloride concentration of the surrounding concrete is still higher than the critical chloride concentration for corrosion to initiate. This leads to the assumption that the protective oxide layer is destroyed causing the corrosion propagation to occur immediately after the replacement, effectively skipping the corrosion initiation time. The 'agent' can decide to perform an inspection and/or maintenance action for each component at each decision step t [year].

In figure 4.1, the structure is shown schematically. The figure includes the identified components (CS1-CS5), the geometrical configuration, and the loading situation.

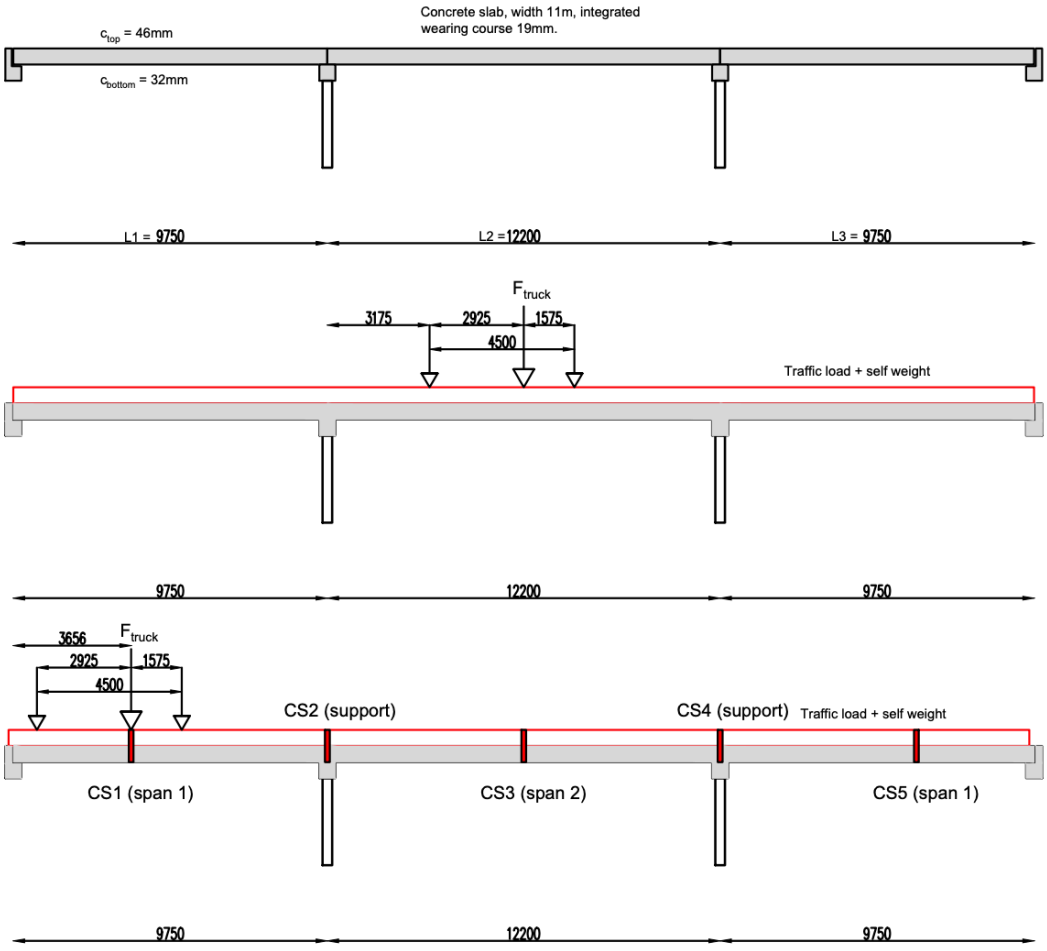


Figure 4.1: Schematics of the case study

4.2. Modelling specifications

In this section, the different modeling specifications used in the calculations are introduced and explained. The section consists of general assumptions for the structure, specifications about the deterioration mechanisms, and the climate scenarios. The input parameters for the calculations are given in tables at the end of the section.

4.2.1. Assumptions

Below some general assumptions are given which are used to set up the case study.

- *Goal setting:*
  The client has expressed that the probability of failure over the lifetime of the structure is not allowed to exceed 1% at any point in its lifetime. The structure must remain operational from the start (2020) to the end of the service lifetime (2050). The inspection and maintenance policy must have lower life-cycle costs than both the condition-based and time-based benchmarks used by the client for other bridges in his care.

- *Probability of failure:*
  The five critical cross-sections are assumed to be statistically independent. The deterioration

of each cross-section, the reinforcement covers, and the reinforcement areas are therefore assumed to be uncorrelated. The probability of failure P_f is assessed using the probability of failure of a component p_i. The structure is a series system with independent components which leads to the following failure probability: $P_f = 1 - \prod_{i=0}^{5}(1 - p_i)$.

- *Loading:*
  The structure is loaded using a combination of a dead load (self-weight), a general traffic load, and a truckload with a dynamic amplification factor. The loading is in compliance with Eurocode 1 part 2: Traffic loads on bridges [57]. The loading is translated towards a critical bending moment for each component $M_{s,i}$ to be used in the limit state. More information about the loading can be found in appendix A.

- *Inspection and maintenance actions:*
  The costs of different actions are related to the cost of a major repair. A minor repair C_m is 70% of performing a major repair. An inspection costs C_i is 1% of the costs of a major repair. The cost of failure C_F and cost of spalling C_s are artificially set higher to include unquantifiable costs, such as the social impact and unavailability of the structure if such an event happens. The ratios can be seen in table 4.5. The setup of the costs depends greatly on the project. For this case study the costs are related to the cost of performing a major repair and are calibrated through the benchmarks. The calibration process is explained in section 5.3.2.

4.2.2. Corrosion of the rebar

The longitudinal reinforcement bars in the concrete slab are subjected to chloride-induced corrosion propagation. The corrosion rate is temperature-dependent and time-invariant, see equation 2.2. The concrete compressive strength is assumed to be time-independent. The corrosion process is independent for each component and described by a gamma process, see section 2.2.3. The gamma process is calibrated by a random value deterioration model, see section 2.2.2. The RVD model considers a combination of uniform and pitting corrosion, see equation 2.13. Input parameters for the RVD model follow from stochastic distributions of existing literature and are summarized in table 4.3. The corrosion rate only changes in time due to temperature, other variables such as relative humidity and chloride concentrations fall outside of the scope. A minor repair action will reset the deterioration rate τ to the initial deterioration rate τ_0 at $t = 0$. A major repair action will reset both the deterioration rate and damage state of a component to the initial values, $d = d_0, \tau = \tau_0$. The bridge is subjected to deicing salts and spans over water, therefore modeling parameters of the corrosion process are chosen similar to structures subjected to marine environments/splash zones in accordance with DuraCrete [24]. As introduced earlier, no corrosion initiation phase is considered for the initial structure or when a maintenance action is performed. The failure probability is used to compute a failure cost C_F, related to the costs of replacement C_r, using a ratio of $a_r = 10000$ in the following manner:

$$C_F = P_F \cdot C_r \cdot a_r \tag{4.1}$$

4.2.3. Spalling of the concrete

The corrosion process causes cracking and eventually spalling to occur due to the internal stresses produced by the expansive corrosion products. The spalling of the component affects the serviceability of the structure and can cause unsafe situations to arise and the corrosion process to accelerate. The spalling process is described in section 2.2.2. The probability of spalling $P_s = P(t \leq t_{spal})$ to occur is calculated using the RVD model to find the probability of spalling for each year $P_{s,CS}$ for each temperature scenario CS. At each decision step t, the probability of spalling is used to calculate the cost of spalling C_s. In table 4.4 the input parameters for the spalling process are described.

$$C_s = P_s \cdot C_r \cdot a_s \tag{4.2}$$

4.2.4. Climate scenarios

Changes in the deteriorating environment have been simulated by temperature scenarios presented by the KNMI [13], see section 2.1.1. Other parameters such as changes in the relative humidity, amount of precipitation and amount of frost days per year can have significant impacts on the corrosion process, however, fall outside of the scope of the research. The effects of climate change on the deteriorating

structure are site-dependent and must follow from the risk analysis performed beforehand.

The temperature scenarios change the corrosion rate in the RVD model, leading to different gamma processes for each temperature scenario. The temperature scenarios are described using key-values for mid-century T_{50} (2050) and end-of-century T_{85} (2085). The key values are accompanied by a 30-year natural variation σ_{30y}, identical for each temperature scenario. The temperature scenarios are also accompanied by a yearly variation σ_y to describe the annual variation of the climate. In table 4.6, the input parameters for the temperature projections are described. More information about climate modeling can be found in chapter 5.

Parameter	Value	Unit
$height$	532	[mm]
$width$	11	[m]
l_1	9.75	[m]
l_2	12.2	[m]
l_3	9.75	[m]
$A_{stnom,span1}$	$\varnothing 20 - 140$	[mm]
$A_{stnom,span2}$	$\varnothing 20 - 160$	[mm]
$A_{stnom,support}$	$\varnothing 20 - 105$	[mm]
$c_{top,nom}$	46	[mm]
$c_{bottom,nom}$	32	[mm]
d_{nom}	$height - c - 1/2d_r - 19$	[mm]
G_{nom}	25	$[kN/m^3]$

Table 4.1: Geometry of the case study

Parameter	Mean	COV	Unit	Distribution	Reference
$depth$	$d_{nom} + 8$	$\sigma = 11.9$	[mm]	Normal	[59]
c_{top}	$c_{tnom} + 19.8$	$\sigma = 19.8$	[mm]	Normal	[59]
c_{bottom}	$c_{bnom} + 8.6$	$\sigma = 10.4$	[mm]	Normal	[59]

Table 4.2: Input parameters for the geometry of the case study

Parameter	Mean	COV	Unit	Distribution	Reference
$A_{st} - 0.91A_{stnom}$	$0.10A_{stnom}$	0.04	$[mm^2]$	Lognormal	[59]
f_{sy}	490	0.10	$[N/mm^2]$	Lognormal	[30]
$i_{corr,20}$	2.586	$\sigma = 1.724$	$[\mu A/cm^2]$	Lognormal	[24]
R	[4-8]	-	-	Uniform	[30]

Table 4.3: Input parameters for the deterioration of the reinforcement

Parameter	Mean	COV	Unit	Distribution	Reference
$d_{r,0}$	25		[mm]	Discrete	
f_c	26.2	0.18	$[N/mm^2]$	Lognormal	[30]
fct	$0.69\sqrt{f_c}$	0.20	$[N/mm^2]$	Normal	[59]
$i_{corr,exp}$	100		$[\mu A/cm^2]$	Discrete	[6]
k_c	1.0		[-]	Discrete	[17]
w	1.0		[mm]	Discrete	[6]
$ME(r_{crack})$	1.04	0.09	[-]	Normal	[6]
f_t	$0.53\sqrt{f_c}$	0.13	$[N/mm^2]$	Normal	[59]

Table 4.4: Input parameters for the spalling of the concrete

Parameter	Description	Ratio
C_r	Major repair	$1 \cdot C_r$
C_m	Minor repair	$0.70 \cdot C_r$
C_i	Inspection costs	$0.01 \cdot C_r$
C_F	Cost of global failure	$P_F \cdot C_r \cdot 10000$
C_s	Cost of spalling per component	$P_s \cdot C_r \cdot 0.2$

Table 4.5: Input parameters for the inspection and maintenance actions

Climate scenario		**GL**	**GH**	**WL**	**WH**	**GL**	**GH**	**WL**	**WH**	
Worldwide temperature change [°C]		+1	+1	+2	+2	+1.5	+1.5	+3.5	+3.5	
Change in airflow pattern		Low	High	Low	High	Low	High	Low	High	Natural variation per 30y
Period	Now (1981-2010)	2050 (2036-2065)				2085 (2071-2100)				
Temperature (°C)	10.1	+1.0	+1.4	+2.0	+2.3	+1.3	+1.7	+3.3	+3.7	± 0.16

Table 4.6: Input parameters for the temperature projections [13]

5 Analysis

This chapter will apply the framework to the case study presented in the previous chapter. The sections are divided into the same steps as for the framework: set up of the physical system, set up of the POMDP, and finding the optimal policy. The previous chapter has already discussed the physical system setup and will be finished in section 5.1. In the first section, the boundaries and initial conditions of the system will be given. In section 5.2, the calculation of all the required components of the POMDP will be performed and explained. An optimal sequence of inspection and maintenance actions for the POMDP is sought by applying a DRL algorithm in section 5.3. The solution will be verified by the time- and condition-based benchmarks. This chapter will discuss the results, and the framework will be addressed in chapter 6.

5.1. Setup of the physical model

Before performing any calculations, the physical system has to be described in detail so no more information is required when starting the analysis. This includes setting boundary conditions, initial conditions, and data collection necessary for the calculations performed in the following sections. Most of the subsections below summarize the data provided in the previous chapter and the description of the case study. The result of this section is a sufficiently defined case study to be used in the setup of the POMDP.

5.1.1. Definition of the goal

The inspection and maintenance plan's goal is described on three requirements: functional, economic, and time. The requirements form boundary conditions for the optimal policy to comply with.

In the previous chapter, the goal of the 'client' is introduced and can be formulated towards a goal in the following fashion:

- Find an optimal sequence of inspection and maintenance actions considering a changing climate where the failure probability does not exceed 1% for each decision step and where the costs are minimized and lower than the benchmark policies for 2020 to 2050.

After finding the optimal policy, the goal will be used to validate that the inspection and maintenance plan complies with the user's wishes.

5.1.2. Climate scenarios

The goal provides requirements for the framework to comply with. The first step after setting the goal would be to perform a risk analysis to identify the critical components, determine the effect of climate change on the structure and determine the governing deterioration mechanism. From the illustrative risk analysis, it has followed that the critical components are the lateral reinforcement subjected to the highest bending moments. The structure is deteriorating due to chloride-induced corrosion, and

the effect of climate change is best described by a temperature-dependent corrosion rate affecting the structure's resistance. Temperature scenarios simulate an accelerated corrosion rate due to climate change.

The temperature scenarios are based on information provided by the Royal Netherlands Meteorological Institute (KNMI) in their climate scenarios '14 [13] and scientific background report [14], described in section 2.1.1. The data from the temperature scenarios with the natural variation averaged over 30 years is described in table 5.1 [13]. The temperature scenarios considered for the case study are the four scenarios presented in table 5.1 and a traditional temperature scenario with a constant average annual temperature.

Climate scenario		**GL**	**GH**	**WL**	**WH**	**GL**	**GH**	**WL**	**WH**	
Worldwide temperature change [°C]		+1	+1	+2	+2	+1.5	+1.5	+3.5	+3.5	
Change in airflow pattern		Low	High	Low	High	Low	High	Low	High	Natural variation per 30y
Period	Now (1981-2010)	2050 (2036-2065)				2085 (2071-2100)				
Temperature (°C)	10.1	+1.0	+1.4	+2.0	+2.3	+1.3	+1.7	+3.3	+3.7	± 0.16

Table 5.1: Annual average temperature for the climate scenarios in the Netherlands

The 30-year averaged natural variation coincides with a 90% confidence interval. The scientific background report [14] describes the natural variability by a Gaussian distribution where the 90% interval corresponds with $\pm 1.645\sigma$ for a 30-year period, with σ being the standard deviation.

Yearly average temperatures are not only subject to variation due to climate change but also due to annual variations. To find the annual temperature variation, the following approach has been taken.

First, a data set was found that gives the annual average temperatures and the expected trend due to global warming [60] from 1952 to 2017.
Next, the trend of climate change was subtracted from the annual average temperatures to obtain the deviation of the expected yearly average temperatures. Afterward, the resulting deviation was fitted to a normal distribution with a mean of zero to get the standard deviation. This method has been applied to both the entire data set and the data set ordered by decade to compare the effects of climate change on the spread of the standard deviation. There appears to be no significant change between the standard deviation of the entire data set ($\sigma_y = 0.374°C$) and the average of the standard deviation over the decades ($\sigma_y = 0.378°C$). This leads to the assumption that the standard deviation of the yearly variation remains constant and identical for each temperature scenario. $\sigma_y = 0.374°C$ is used to simulate the annual variations.

The approaches described below are adopted to incorporate both the variability due to yearly variations σ_y and the 30-year natural variation of the temperature projections σ_{30y}.

Different approaches have been taken to incorporate the variability of the temperature for individual components of the model. The parts where the temperature scenarios were used are: (i) *Transition model and probability of spalling* $P(s'|s,a)$ for the POMDP and construction of the CDF for the spalling time $P(T_{spal} \leq t)$, (ii) To update the belief of the agent over the different temperature scenarios. (iii) Simulating different temperature trajectories as training data for the DRL agent. The different approaches are described below.

(i) *Transition model and probability of spalling:* The temperature-dependent transition models $P_j(s'|s,a)$ and cumulative distribution functions $P_j(T_{spal} \leq t)$ are used in combination with the belief-space $b(s_t)$ to compute the updates, more on that in the section 5.2. For each scenario, a temperature projection is sampled from the mid-century distribution $T_{50} \sim N(\mu_{50}, \sigma_{30y})$, and the end of century distribution $T_{85} \sim N(\mu_{85}, 2 \cdot \sigma_{30y})$. This is done to stay in line with the temperature scenarios given by the KNMI, which do not include a yearly variation. From the two temperatures, a bilinear temperature projection is created:

$$T_{cp}(t) = \begin{cases} \frac{T_{50}}{2050-2020} \cdot (t - 2020) & \text{if } 2020 \leq t < 2050 \\ \frac{T_{85}-T_{50}}{2085-2050}(t - 2050) + T_{50} & \text{if } 2050 \leq t \leq 2085 \end{cases} \tag{5.1}$$

In figure 5.1a, such a bi-linear model can be seen. In the figure, the colored bands correspond to the 90% confidence interval

(ii) *belief state over the climate scenarios:* The belief over the climate scenarios is updated at each decision step. To update the belief, the probability of a temperature occurring for a given climate scenario is assessed using the previous year's temperature as evidence. This belief state is updated using Bayesian Updating described in section 2.3.1. To update the belief, the probability of occurrence of a temperature T for a given temperature scenario is required. The exact process of updating the belief is provided in section 5.2.

$$P(T_{cs}(t) = T(t)) \text{ with: } T_{cs}(t) \sim N(\mu_{cs}(t), \sigma_y) \tag{5.2}$$

where: $P(T_{cs}(t) = T(t))$ is the probability of observing temperature $T(t)$ for a random variable describing the temperature for a climate scenario $T_{cs}(t)$. The RV $T_{cs}(t)$ is described by a normal distribution with mean $\mu_{cs}(t)$ and a yearly standard deviation σ_y.

The natural variation of the climate scenarios σ_{30y} is neglected due to the difficulty of assessing the probability of occurrence of such an event when the mean is an RV itself (see equation below). This assumption is grounded since the standard deviation of the 30-year natural variation ($\sigma_{30y} = 0.1$) is small compared to the yearly standard deviation ($\sigma_y = 0.374$).

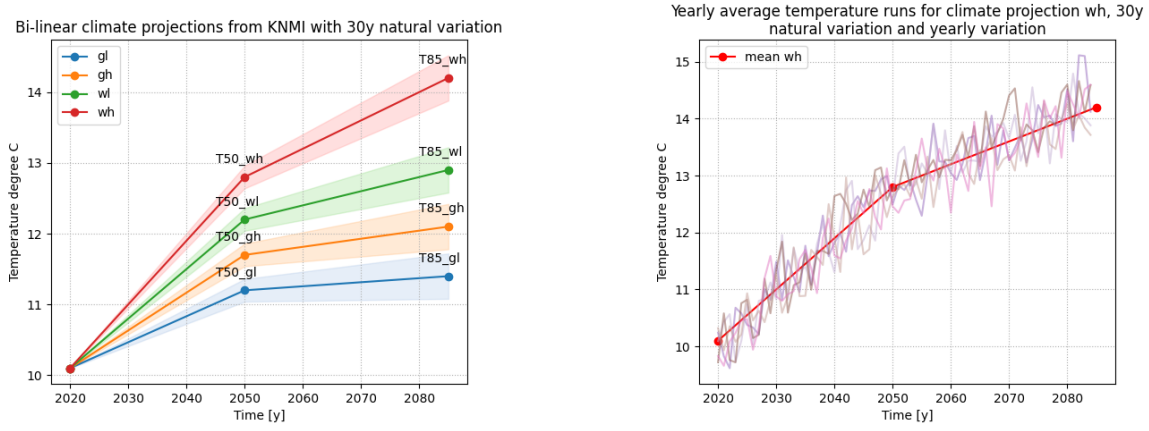
$$T_{CS}(t) \sim N\left(M_{CS}(t), \sigma_y\right) \text{ with: } M_{CS}(t) \sim N(\mu_{cs}(t), \sigma_{30y})$$

where: $M_{CS}(t)$ represents the mean value as a random variable.

(iii) *Training data:* The simulations used to train the model apply both the 30-year natural variation and the yearly variation where the choice over which temperature scenario is followed is picked uniformly over the possible scenarios. This is done to make the model robust and closer to reality. Simulations are sampled in the following form:

(a) Sample T_{50} and T_{85} to create the bi-linear temperature function given in equation 5.1.
(b) Use the bilinear function as a mean $\mu_{cp}(t) = T_{cp}(t)$ for a Gaussian distribution with a yearly variation: $T \sim N(T_{cp}(t), \sigma_y)$

Figure 5.1b shows an example of ten sampled temperature runs for climate projection 'WH.' The straight line corresponds to the mean annual average temperature trend from table 5.1.



(a) KNMI bilinear climate projections with 90% confidence interval

(b) Climate projection 'WH', random temperature runs

Figure 5.1: Simulated climate scenarios

5.1.3. Stochastic deterioration process

The deterioration of the structure is described by the propagation of corrosion due to the ingress of chloride. The corrosion process has initiated at the start of the inspection and maintenance plan. To set up the stochastic deterioration process, the following two models need to be set up. First, a random value deterioration model is set up for each temperature scenario using a combination of pitting and uniform corrosion. Next, the RVD model calibrates an SPD model as a gamma process for each temperature scenario.

The RVD model uses a temperature-dependent corrosion rate as presented in equation 2.2. The residual cross-sectional area of the reinforcement steel $A_s(t)$ is used as presented in equation 2.13. The expected damage at the end of service life is used to calibrate the gamma process. The amount of cross-section loss represents the damage of the cross-section compared to the initial cross-section; see equation 5.3.

$$d_j(t) = \frac{A_s(0) - A_{s,j}(t)}{A_s(0)} \tag{5.3}$$

where: $d_j(t)$ represents the damage of the cross-section for climate scenario j, $A_s(0)$ represents the initial cross-sectional area of the rebar and $A_{s,j}(t)$ is the residual cross-sectional area of the rebar for climate scenario j at time-step t.

The RVD model, as described in section 2.2.2, combines pitting and uniform corrosion. In the gamma process, it is assumed that corrosion only occurs uniformly. Since the resistance of the rebar only depends on the residual cross-sectional area and not on the geometry of the bar, this assumption is grounded.

A damage increment is sampled for each component for a temperature scenario at each time step according to the equation below.

$$\Delta d_{i,j}(\tau_i) \sim Ga(v(\tau_i) - v(\tau_i), u_j)$$

where: $\Delta d_{i,j}(\tau_i)$ represents the damage increment for component i under climate scenario j, τ_i represents the deterioration rate for component i, $v(\tau_i)$ represents the shape parameter of the component and u_j is the scale parameter which depends on the climate scenario.

The scale parameter u_j and shape parameter $v(\tau_i)$ need to be calibrated to set up the gamma process. The calibration is based on the expected damage from the RVD model. A coefficient of variation (COV) of 20% is assumed for the gamma process.

$$u_j = \frac{\left(\mathbb{E}(d_j(30)) \cdot COV\right)^2}{\mathbb{E}(d_j(30))} \tag{5.4}$$

where: $\mathbb{E}(d_j(30))$ is the expected damage at yat each time-step ear 30 according to the RVD model for climate scenario j.

$$a = \frac{COV^2}{30^\beta}$$

$$v(\tau_i) = a \cdot \tau_i^\beta$$

where: β is found by fitting a power-law function to the RVD model and depends on environmental conditions, corrosion causes, material properties, etc. [8]. A value of 1.22 is found for this study.

The RVD model describes the random variable $A_s(t)$. The model is analyzed using a crude Monte-Carlo analysis by drawing 100,000 samples in Python of the stochastic distributions given in tables 4.2 and 4.3 for the initial values. Next, the degradation is progressed over time according to equation 2.13 to find the progress of $A_s(t)$ over time. This setup is repeated for each of the climate scenarios. To form the SPD model, the remaining average cross-section at the end of the planning horizon is required and is described in table 5.2. The table displays $A_s(t = 30y)$ according to the mean of the samples and the 95% upper- and lower bound and a translation towards the damage of the cross-section.

Scenario	average [mm2]	lower bound [mm2]	upper bound [mm2]	damage
gl	251.1	97.15	302.1	0.201
gh	250.5	94.88	302	0.203
wl	249.9	92.59	301.9	0.205
wh	249.1	89.65	301.8	0.207
trad	252.5	102.3	302.4	0.196
Initial	314.16	-	-	0

Table 5.2: Description of the random variable $A_s(t = 30y)$ for the different climate scenarios

The SPD model uses the damage of each climate scenario given in table 5.2 to set up a gamma-process for each climate scenario. The gamma process is required to simulate a process of random positive-damage increments per component and to update the belief state according to a transition model. In table 5.3, the damage according to a gamma process following a climate scenario is described according to 10,000 random runs per climate scenario by the average and a lower- and upper bound of a 95 % confidence interval.

Scenario	Variable	average	lower bound	upper bound
gl	Damage	0.207	0.138	0.291
	Area [mm2]	249.1	222.7	270.8
gh	Damage	0.208	0.139	0.292
	Area [mm2]	248.8	222.4	270.5
wl	Damage	0.210	0.139	0.296
	Area [mm2]	248.2	221.2	270.5
wh	Damage	0.212	0.142	0.299
	Area [mm2]	247.6	220.2	269.5
trad	Damage	0.201	0.135	0.284
	Area [mm2]	251.0	271.7	224.9

Table 5.3: Description of the variable $d_j(t = 30y)$ for the different climate scenarios.

In figure 5.3, the cross-sectional damage $d_j(t)$ is plotted against time for the RVD-model and SPD-model for each temperature scenario. The plot consists of twenty random simulations and the mean of all simulations for each combination of the deterioration model and temperature scenario.

It can be observed that the mean of the RVD model and SPD model are similar. This observation is also verified by the damage at the end of the planning horizon being identical between both the RVD-model (table 5.2) and the SPD-model (table 5.3). The SPD model also takes random positive-damage increments, taking in the temporal variability inherent to the degradation process [35].

Another observation is that the RVD model's variability is larger than the SPD model's. The origin of the variability is due to the high variance of the random variable corresponding to the reference corrosion current density $i_{corr,20}$, which has a COV of 0.67. At the same time, the SPD model has been set up with a COV of 0.20. This assumption has been taken since the gamma process gives illogical results when applying a relatively large covariance. Equation 2.24 describes how the gamma distribution's mean value and standard deviation can be calculated. Below, this relation has been applied to the definition of the coefficient of variation to find a link to the shape function.

$$COV = \frac{\sigma}{d_m} = \frac{1}{\sqrt{v(\tau_i) - v(\tau_i - 1)}} \rightarrow v(\tau_i) - v(\tau_i - 1) = \frac{1}{COV^2}$$

For a high coefficient of variation, the difference between the shape parameters $v(\tau_i) - v(\tau_i - 1)$ becomes smaller, which requires a large scale parameter to compensate for this slight difference. This results in the SPD model taking either small or large positive damage increments, which leads to illogical results. To further illustrate this, the same process has been initialized with a higher covariance in figure 5.2. Since this assumption has been applied to each climate scenario, it will not affect the framework's validity. It will therefore be used in the continuation of this study.

It can also be observed that the differences between cross-sectional damage in each temperature

scenario are minimal. The minimal difference between the scenarios is due to how the climate scenarios are included in the limit state. The climate scenarios are solely included in the temperature-dependent corrosion rate given in equation 2.2. The temperature-dependent corrosion rate applies a factor K_T to represent the effect of temperature change on the corrosion rate. Since K_T and the planning horizon are relatively small, this leads to minor differences between the scenarios. This observation must be noted since this might lead to the policy from the framework taking no different inspection and maintenance planning for different climate scenarios. However, the effect of climate change is present in many risk categories [3], is case-specific, and might therefore be more extreme. However, the case study still shows an example of how the framework can be applied and is sufficient to continue for now.

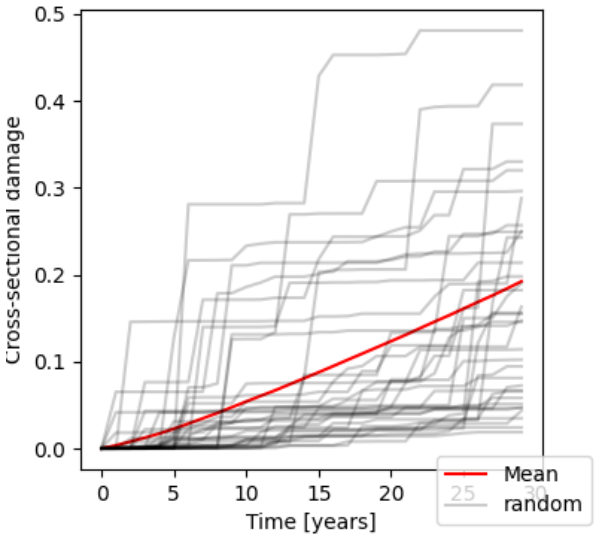


Figure 5.2: Gamma process for high COV

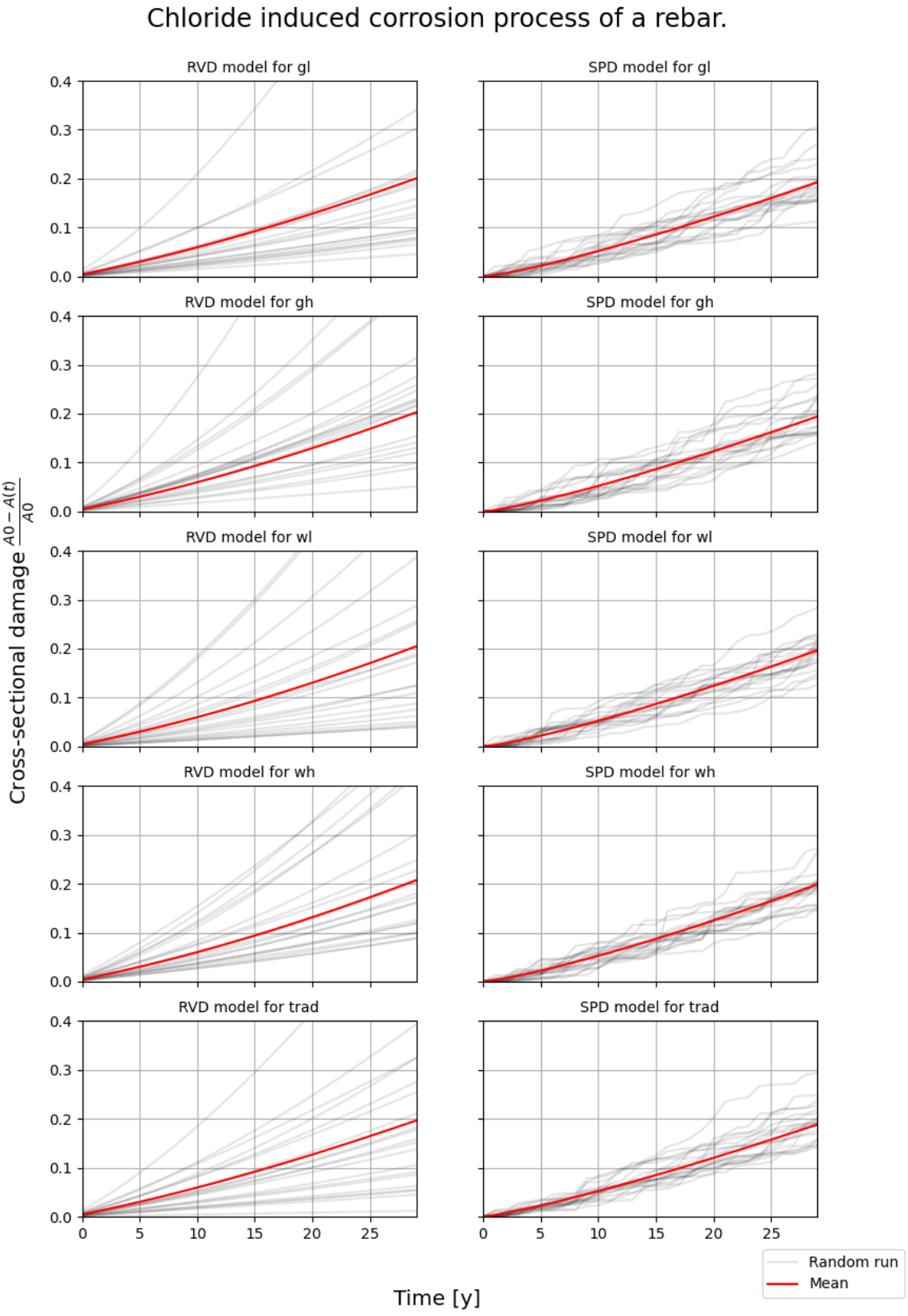


Figure 5.3: Chloride induced corrosion process for Random Variable Degradation model (RVD) (left) and Stochastic-Process Degradation model (SPD) (right) for the different climate scenarios.

5.1.4. Concrete spalling

The different temperature scenarios enter the equation for spalling through the corrosion current density $i_{corr}(t)$ as described in equation 2.2. Spalling also depends on the height of the concrete cover c, which is different for the bottom and top of the slap. This requires ten distinct cumulative distribution functions to cover all combinations of temperature scenarios and concrete cover height.

Section 2.2.2 under chloride-induced corrosion propagation explains how concrete spalling can be linked to different temperature scenarios. This is done by combining equations 2.20 and 2.21 to form the following equality:

$$\frac{x_{cr}}{11.6} + 0.0114 k_R \left(\frac{w - 0.05}{k_c ME(r_{crack}) r_{crack}} \right) = \int_{t_0}^{t_{sp}} i_{corr}(t) dt$$

The left side of the equality sign represents the resistance of the cover against spalling. The right side of the equality sign represents the loading or degradation of the concrete slap. The limit state in equation 5.5 is used to compute the cumulative distribution function.

$$R_2 = \frac{x_{cr}}{11.6} + 0.0114 k_R \left(\frac{w - 0.05}{k_c ME(r_{crack}) r_{crack}} \right)$$

$$S_2 = \int_{t_0}^{t_{sp}} i_{corr}(t) dt$$

$$Z_2 = R_2 - S_2 \tag{5.5}$$

The cumulative distribution function is based on the random variable T_{spall}. The limit state consists of combinations of random variables described in table 4.4. For each combination of climate scenario and concrete cover height, 100,000 samples were drawn from stochastic distributions given in table 4.4 and checked for which time-step the loading exceeds the resistance. The time-step when the loading is larger than the resistance represents the time to spalling $t_{sp} \in T_{spall}$. By performing many simulations, the random variable T_{spall} is sufficiently described, and the probability of spalling occurring for a certain time-step can be found. In table 5.4, the random variable T_{spall} is described according to the mean value and a lower- and upper bound for a 95% confidence interval.

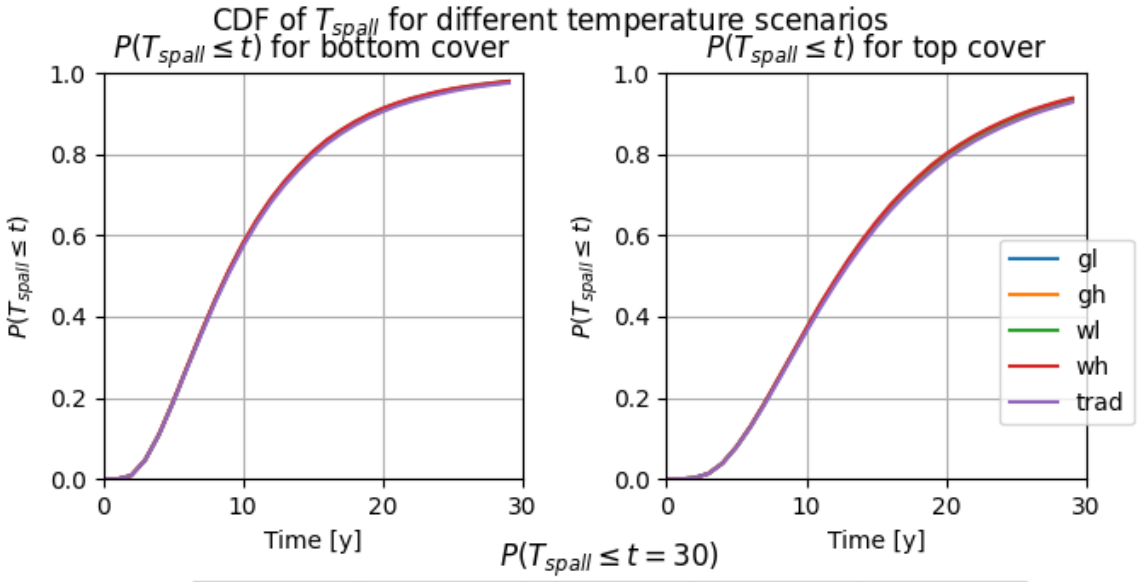
	Top cover			*Bottom cover*		
Scenario	**Average**	**Lower bound**	**Upper bound**	**Average**	**Lower bound**	**Upper bound**
gl	13.9	3	36	10.1	2	28
gh	13.8	3	36	10.0	2	28
wl	13.7	3	36	9.98	2	28
wh	13.6	3	35	9.94	2	27
trad	14.0	3	37	10.2	2	28

Table 5.4: Description of the random variable T_{spall} for the different temperature scenarios and concrete cover heights.

Figure 5.4 shows the cumulative distribution function (see equation 5.6) for the different climate scenarios and concrete covers. Underneath the figure, a table is given, which gives the probability that spalling has occurred at the end of the lifespan if no maintenance is performed.

$$F_{T_{spall}}(t) = P(T_{spall} \leq t) \tag{5.6}$$

where: T_{spall} is the random variable time to spalling, which depends on the concrete cover and climate scenario, $F_{T_{spall}}(t)$ is the cumulative distribution function of T_{spall} and t represents the time-step.



$P(T_{spall} \leq t = 30)$

	gl	gh	wl	wh	trad
Bottom	0.97695	0.9778	0.97846	0.97939	0.97515
Top	0.93195	0.93388	0.93577	0.93784	0.92773

Figure 5.4: Cumulative distribution function for spalling under the temperature scenarios for the bottom concrete cover (left) and top concrete cover (right)

The following observations can be made from the description of the random variable T_{spall} in table 5.4. First, there is an approximately three-year difference between the time of spalling for the top cover and the bottom cover. This can lead to a different sequence of inspection- and maintenance actions for components with either a bottom- or top surface. Second, there is quite a big spread between the lower- and upper bound of the confidence interval. This spread can be explained due to the high variability of the RVD model. Third, there is barely any difference between the average time for spalling between the different climate scenarios. Again, this might result in the framework not reaching its full potential, as explained in the previous subsection. This observation is also supported by figure 5.4, where the different CDFs can barely be distinguished.

5.1.5. Input data

For the physical modeling, the stochastic variables must be identified with experiments or existing literature. In this research, values of the existing literature are used. With the input data, probability values can be derived.

Stochastic deterioration process

The stochastic deterioration process consists of the random value degradation model and the stochastic degradation process. For the random value degradation model, the following parameters are required:

- Initial reinforcement area $A_{st}(0)$
- Yield strength of the reinforcement f_{sy}
- Reference corrosion current density $i_{corr,20}$
- Initial reinforcement diameter $d_{r,0}$
- Conversion factor of uniform to pitting corrosion R

The stochastic distribution, mean, and covariance is given for each parameter with the corresponding reference in table 4.3.

To set up the stochastic deterioration process, the following assumptions have been taken:

- Each component has its damage process and is assumed to be independent of one another.

- No corrosion has yet occurred at any component and is assumed to have been initiated at the start of the inspection and maintenance plan at $t = 0$.
- Only temperature is assumed to impact the corrosion rate; no other climate effects are considered.

Concrete spalling

The spalling of the concrete is modeled with the random value degradation model using a temperature-dependent corrosion rate. From the limit state function, it follows that the following parameters are required:

- Thickness of the concrete cover c
- The concrete compressive strength f_c
- The splitting tensile strength of the concrete f_{ct}
- Correction factor for the rate of loading k_R
- Confinement factor k_c
- Model error $ME(r_{crack})$
- Tensile strength of the concrete f_t
- Crack width limit w

The stochastic distribution, mean, and covariance is given for each parameter with the corresponding reference in table 4.4.

To calculate the limit state function Z_2, the following modeling assumptions have been taken:

- The deterioration rate τ is assumed to be directly proportional to the time t. This assumption requires that only full resets of the deterioration rate can happen if a maintenance action is performed.
- The starting diameter $d_{r,0}$ is assumed to be constant. After a maintenance action is performed and the deterioration rate τ is reset, the starting diameter would be smaller than the initial diameter. This would lead to different values for the probability of spalling. It is, however, assumed that such maintenance actions do not often occur during the structure's lifetime and that this effect is negligible to keep direct proportionality between the deterioration rate τ and the time step t.
- No cracking has yet occurred for each component. Process assumed to initiate at the start of the inspection and maintenance plan.

5.1.6. Inspection and maintenance actions

The final step in translating the physical model is to list the costs and consequences of each action that can be taken at each decision step t. The actions consist of performing a maintenance and/or inspection action; the combinations considered per component are given in table 5.5. The setup of the costs is described in table 4.5, with an illustrative major repair cost of $10,000€.

Action	Description	Consequence	Cost
Action 1, a_1	Do nothing, no inspection	Transition according to transition model	0
Action 2, a_2	Do nothing, inspection	Transition according to transition model and update belief according to observation model	C_i
Action 3, a_3	Minor repair, Inspection	Transition according to transition model, reset the deterioration rate τ and update belief according to observation model	C_m
Action 4, a_4	Major repair, no inspection	Reset the deterioration rate τ and the damage of the component d_i to the initial state.	C_r

Table 5.5: Inspection and maintenance actions

5.2. Setup of the POMDP

The physical model has been set up and needs to be translated towards a POMDP, the mathematical framework of the sequential decision-making problem. As explained in section 2.3.1, the POMDP is described according to the tuple: $< S, A, R, O, Z, T, \gamma >$.

In the first subsection, the components of the POMDP will be introduced and the state-space S will be discretized in order to significantly decrease the computational time required. The effect of the actions A on the state S and the rewards R will be given. In the second subsection, the SPD model will be translated towards a transition model T. In the final subsection, the computation of the failure probability will be described, which is required for the reward function for R. After these steps, the POMDP is sufficiently described to start finding an optimal inspection and maintenance policy.

5.2.1. Framework POMDP

The state of the environment is described as variables for the components and variables of the system. Each component i is described based on the deterioration rate of the component τ_i and the damage of the component D_i. The age is directly proportional to the deterioration rate and is assumed to be known at every decision step. The damage of a component is discretized in ten evenly spaced states from no damage up to 20% damage. The system also has a variable CS which tracks the belief over the possible climate scenarios. The climate scenarios considered are the ones provided by the KNMI. The definition of each of the different variables is given below.

$$D_i \in [0\%, 2.22\%, 4.44\%, ..., 20\%]$$

$$\tau_i \in [0, 1, ..., 29]$$

$$CS \in [\prime GL\prime, \prime GH\prime, \prime WL\prime, \prime WH\prime, \prime TRAD\prime]$$

The entire state of the environment S is defined by the combination of each of the sets and is given below.

$$S = [[D_1, D_2, D_3, D_4, D_5], [\tau_1, \tau_2, \tau_3, \tau_4, \tau_5], CS]$$

At each time step t, the hidden state of the environment s_t is a subset of S. The POMDP applies a probability distribution over the state-space S to represent a belief state **b**. The probabilities over the hidden-state $b(s)$, for all $s \in S$, form the belief vector **b**.

The four I&M actions considered for the case study are given in table 5.5. The costs for each action are described in table 4.5. An inspection a_2, causes the belief state to be updated according to the observation model O. For this case study, the observation model inspects the correct damage state of a component with probability p and the adjacent states with probability $(1-p)/2$ (see equation 3.7). The probability of observing the correct state p has a value of 90% accuracy.

A minor repair a_3 consists of the deterioration rate of a component to be reset to zero $\tau_i \leftarrow 0$. This effectively causes the corrosion process to reset which slows the progress of corrosion while maintaining the same damage state. It has also been assumed that an update of the belief state is performed according to the same observation model O since a repair causes the reinforcement to be examined.

A major repair a_4 causes the entire corrosion process to be reset, including the damage state, effectively replacing the component. An example of such a transition model can be seen in equation 3.6.

The reward function R consists of a combination of risk-costs C_{risk} and action-costs C_{action}, as explained in section 3.2.4.

5.2.2. Degradation process

The degradation process is described in the previous section and needs to be translated towards a transition model $T = p(s_{t+1}|s_t, a_t)$. The transition model describes how the environment transitions from state s_t towards state s_{t+1} in the succeeding time step according to action a_t. The transition model is required to update the belief state using Bayesian updating, see equation 2.33.

In this subsection, it will be explained how to obtain the transition model for the deteriorating environment given that no maintenance action occurs according to the SPD model described in the previous section.

In order to describe all possible transitions the environment can take, a transition model needs to be constructed for every combination of deterioration rate τ and climate scenario CS. The transition models have been constructed as follows. First, a gamma process was initialized for thirty evenly discretized starting states. The gamma process is identical to the one explained in section 5.1.3. The initial cross-sectional areas vary between +40% ($439mm^2$) and -20% ($251.3mm^2$). Second, 10.000 random runs of each starting state were simulated and state transitions were logged. Third, the transition model was constructed for each deterioration rate according to equation 5.7. And fourth, this process was repeated for each temperature scenario.

$$p(s_{i+1}|s_i) = \frac{\text{number of transitions from } s_i \text{ to } s_{i+1}}{\text{total number of transitions}} \tag{5.7}$$

The results are 10x10 component transition matrices for each deterioration rate and climate scenario, totaling 150 distinct matrices. The values below the diagonal of the matrix are zero since transitions towards a better damage state don't exist. The matrix is diagonal and the rows sum up to one. The first three values from the diagonal are given below for the traditional climate scenario and the worst climate scenario 'wh' for deterioration rates $\tau_0, \tau_{14}, \tau_{29}$ in table 5.6.

	scenario					
	wh			*trad*		
	$p(s_i\|s_i)$	$p(s_{i+1}\|s_i)$	$p(s_{i+2}\|s_i)$	$p(s_i\|s_i)$	$p(s_{i+1}\|s_i)$	$p(s_{i+2}\|s_i)$
τ_0	0.986	0.013	0.001	0.983	0.016	0.001
τ_{14}	0.676	0.299	0.02	0.692	0.29	0.017
τ_{29}	0.623	0.342	0.032	0.653	0.322	0.022

Table 5.6: Transition probabilities for climate scenarios 'wh' and 'traditional' for begin, mid, and end of planning horizon

From the table, the following observations can be made. First, the probability for a component to remain in its current state $p(s_i|s_i)$ becomes smaller for higher deterioration rates. Second, there is a slight increase in the probability to succeed to a worse damage state for the worst climate scenarios compared to the traditional climate scenario. Both observations are in line with what is expected and found in the previous section.

5.2.3. Failure probability

The failure probability is used in the computation of the reward function. In the calculation of the failure probability, it is assumed that the case study can be examined as a series system with independent critical components, as is stated in chapter 4.

The calculation of the initial failure probability and introduction of the limit state is covered in Appendix A. In this section, it will be discussed how conditional failure probabilities are used in combination with the belief state to calculate the failure probability.

The failure probability of a critical component i is described as $P(F_i)$. However, the damage state is hidden and only the belief state is known. The belief state, which is a probability distribution over the possible damage states, can be used to compute a failure probability of a given component using the law of total probability, see the equation below.

$$P(F_i) = \sum_{n=1}^{10} P(F_i|d_n) \cdot P_i(d_n)$$

where: n is the index of the discretized damage states, $P(F_i|d_n)$ is the conditional probability of failure of component i if in damage state d_n and $P_i(d_n)$ is the probability of being in damage state d_n for component i, which follows from the belief state **b**.

In table 5.7, the conditional failure probabilities are given for each combination of distinct critical components and damage states. The conditional failure probabilities are calculated using a Monte-Carlo analysis using importance sampling. 2,000,000 samples are drawn of the stochastic distribution given in tables 4.1, 4.2 and 4.3 per combination of damage state and component.

	Short span	*Long span*	*Supports*
Damage state	**(CC1, CC5)**	**CC3**	**(CC2, CC4)**
d_1	1.18E-07	2.31E-07	1.56E-07
d_2	2.76E-07	4.43E-07	3.69E-07
d_3	6.45E-07	8.98E-07	8.75E-07
d_4	1.65E-06	1.86E-06	2.08E-06
d_5	3.41E-06	4.28E-06	4.96E-06
d_6	6.94E-06	8.81E-06	1.19E-05
d_7	1.44E-05	1.81E-05	2.57E-05
d_8	2.95E-05	3.54E-05	5.71E-05
d_9	6.46E-05	6.90E-05	1.25E-04
d_{10}	1.37E-04	1.40E-04	2.63E-04

Table 5.7: Conditional failure probabilities $P(F_i|d_n)$ for each component and damage state.

With the failure probability according to the belief state of each component known, one can determine the failure probability for a series system with statistically independent components as follows.

$$P_F = 1 - \prod_{i=1}^{5}(1 - P(F_i))$$

where: P_F is the failure probability of the system and i is the index of the critical component.

From table 5.7 the following observations can be made. First, the conditional failure probability of the critical components at the supports (CC2, CC4) is the highest. This observation can lead to the optimal policy performing I&M actions more frequently for those components. Second, the conditional failure probabilities of the different components are quite similar. This observation should lead to the actions being evenly distributed over the components, and therefore the policy should not be too different.

The final observation is regarding the last damage state in the table. The damage state ($d_{10} = 0.20$) corresponds to the probability of failure given that a component is in damage state d_{10}. However, since this is the final damage state, the damage state is actually $d_{10} \geq 0.20$. This difference in formulation means that the $P(F_i|d_{10})$ undershoots the failure probability if a component is in a worse state than d_{10}. This observation requires that policy needs to be checked on whether the agent decides to keep the components in the worst damage state.

5.2.4. Updating of the belief

The belief states $\mathbf{b}_t$ are used to determine the transition model, calculate the probability of failure and spalling, and as input for the agent to determine its policy.

The transition model uses the belief over the climate scenarios $b(CS)$ in order to obtain the transition model by taking the weighted average of the corresponding transition matrices as can be seen below.

$$p_\tau(s'|s) = \sum_{j \in CS} p_{\tau,j}(s'|s) \cdot p(j)$$

where: $p_\tau(s'|s)$ is the transition matrix of deterioration rate τ, CS refers to all possible climate scenarios, $p_{\tau,j}(s'|s)$ is the transition matrix of deterioration rate τ following climate scenario j and $p(j)$ is the probability of being in climate scenario j, which follows from the belief state.

The belief for the succeeding timestep $\mathbf{b}_{t+1}$ for the climate scenarios are updated through the annual average temperature, T_t. Each temperature scenario is described by a time-dependent random variable $T_j(t)$, with j indexing the climate scenario. $T_j(t)$ is described by a gaussian distribution with a

time-dependent mean $\mu_j(t)$ and a constant yearly standard deviation $\sigma_y = 0.374$, $T_j(t) \sim N(\mu_j(t), \sigma_y)$. Updating the belief state over the climate scenarios is done according to equation 5.8. The prior distribution at $t = 0$ is initialized as uniformly distributed over the possible climate scenarios.

$$p(CS = j|T_t) = \frac{p(T_t|CS = j) \cdot p(CS = j)}{p(T_t)} \tag{5.8}$$

where: $p(CS = j|T_t)$ is the posterior distribution, $p(T_t|CS = j)$ is the likelihood of T_t occurring given climate scenario j and $p(T_t) = \sum_{k \in CS} p(T_t|CS = k)p(CS = k)$ is the normalizing constant.

Updating of the belief state is done according to equation 5.10. If no observation occurs, then this equation can be simplified to the form provided in equation 5.9.

$$b(d')_i = \sum_{d \in D} p(d'|d, \tau_i)b(d)_i \tag{5.9}$$

$$b(d')_i = \frac{p(o|d')}{p(o|b(D)_i)} \sum_{d \in D} p(d'|d, \tau_i)b(d)_i \tag{5.10}$$

The update can also be affected by a maintenance action. The minor repair leads to a reset of the deterioration rate τ_i and a major-repair resets both the deterioration rate and damage state of a component. The updating of the belief state is presented in pseudo-code in algorithm 1.

Algorithm 1 belief state update

Initialize belief state
for $t = [1, T]$ **do**
  $b_{t+1}(CS) \leftarrow$ According to equation 5.8
  for $i = [CC1, CC2, CC3, CC4, CC5]$ **do**
    $\tau_i \leftarrow \tau_i + 1$
    if a_i is *do nothing* **then**
      $b(d')_i \leftarrow$ According to equation 5.9
    else if a_i is *inspect* **then**
      $b(d')_i \leftarrow$ According to equation 5.10
    else if a_i is *Minor repair* **then**
      $\tau_i \leftarrow 0$
      $b(d')_i \leftarrow$ According to equation 5.10
    else if a_i is *Major repair* **then**
      Initialize damage state and deterioration rate of component
    end if
  end for
end for

5.3. Finding the optimal policy

In the previous section, each component of the POMDP is properly defined which finalizes the setup of the environment. The next step is to find an optimal inspection and maintenance policy for this environment which will be the topic of this section.

This section covers the applied DRL algorithm, and the setup of the benchmarks, and presents the results found for the optimal policy. Each result is shortly discussed in the corresponding section. Also, an extra analysis is provided which illustratively heightens the effects of temperature change on the corrosion rate to make the effects of climate change more prevalent.

5.3.1. Algorithm

The main categories of DRL algorithms have been explained in section 2.3.3. The distinction lies in whether a value function or the policy itself is parametrized by means of a Deep Neural Network. The

parameterization is required to generalize the experience of the agent to the entire state space and also to significantly lower the required computational effort.

Both categories of algorithms take as input the belief state of the POMDP, the difference lies in the output of the algorithms. Value-based methods, such as DQN, would output an action value for each possible action. Since the amount of possible actions scales exponentially with the number of components, such an algorithm quickly becomes unusable. Due to a large number of possible actions, DQN has not been applied to find an optimal policy. A large number of actions results in an explosion of network parameters to be trained, which requires lots of training time and difficulty to converge toward an optimal policy.

Actor-critic methods alleviate the problem of keeping track of a value for each possible action and directly try to map states to actions. It provides a probability mass distribution over the possible actions as an output of the network. In order to tackle the exponential scaling of the action space with the number of components, DCMAC [8] assumes conditional independence between the actions of the components which leads to linear scaling with the number of components (see equation 5.11). For the case study, the number of components is 5 and the possible actions per component are 4. Without assuming conditional independence between actions of components, the policy function would have an output layer of $4^5 = 1024$ nodes. By applying decentralization, the number of output nodes reduces greatly to $4 * 5 = 20$ nodes.

$$\pi(a|s) = \prod_{i+1}^{5} \pi_i(a_i|s) \tag{5.11}$$

where: $\pi(a|s)$ is the policy of the entire system, i refers to the index of the component and $\pi_i(a_i|s)$ refers to the policy of a single component.

Instead of having one probability mass distribution over all possible combinations of actions, it applies a probability mass distribution per component. For instance, the first four output nodes are the probability mass distribution over the actions for CC1.
DCMAC applies one actor network for the entire system. This means effectively that one DRL agent is employed to make decisions over all components. The actor-network is given the full centralized state space and therefore makes decisions on one component while taking into account the state of every other component.

The algorithm applied to the case study is DCMAC with PPO used for the actor-critic network. The actor-network is parametrized by parameters θ and the critic-network by parameters θ_v. Both the actor- and critic-network are parametrized by the same Deep Neural Network architecture. PPO has been chosen due to the ability to maintain the stability and reliability of trust-region methods while being much simpler in implementation and achieving better performance than, for instance, TRPO. In figure 5.5, an illustration is provided of the input, network, and output of the applied framework. In the figure, the output of the actor-network is five probability mass distributions over the possible actions per component represented by horizontal bars. By sampling from these distributions, the final action-vector **a** is found.

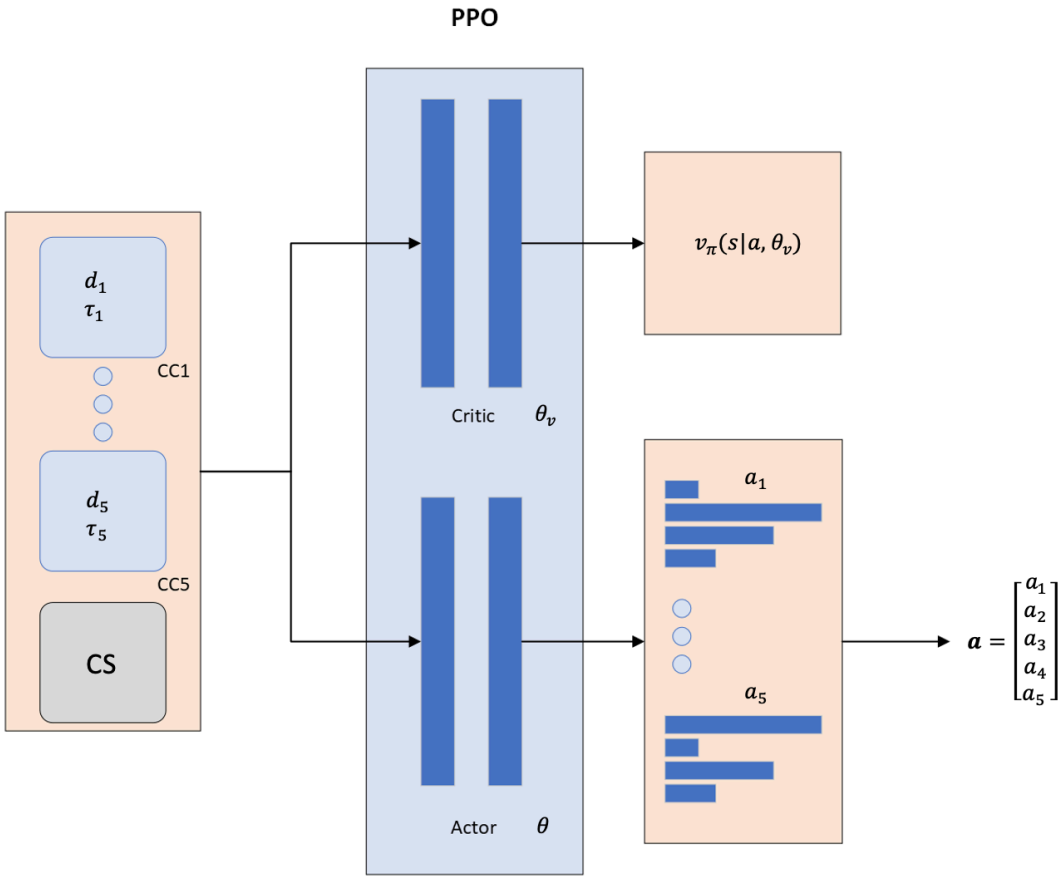


Figure 5.5: DCMAC architecture for case study

The POMDP has been created by using OpenAI's Gym, which is a toolkit for reinforcement learning research [61]. The environment is summarized in algorithm 2. The modeling of the DRL algorithm has been performed using Stable Baselines3 (SB3) [62]. SB3 contains open-source implementations of DRL algorithms in Python. Extensive documentation, examples, and source code can be found at the following location: `https://github.com/DLR-RM/stable-baselines3`. SB3 has PPO as one of the available algorithms and allows for a decentralized policy to be applied. An example of the algorithm which PPO applies to train the model is given in algorithm 3, which follows from documentation provided by [63].

Algorithm 2 Bridge Environment

for $episode \leftarrow [0, N]$ **do**
  $b(s_t) \leftarrow$ reset environment
  $CS \leftarrow pickRandom([' gl',' gh',' wl',' wh',' traditional'])$
  Initialize probabilities P_f, P_s
  done is *False*
  for $t \leftarrow [1, T]$ **do**
    $t \leftarrow t + 1$
    $\pi(a|s, \theta) \leftarrow Actor(b(s_t))$
    $v_\pi(s|\theta_v) \leftarrow Critic(b(s_t))$
    $\mathbf{a} \leftarrow \pi(\mathbf{a}|s, \theta)$
    belief state update (algorithm 1)
    $C_{action} \leftarrow \sum_{i=1}^{5} C_{action,i}$
    $C_s \leftarrow \sum_{i=1}^{5} P_{s,i} * C_r * 0.20$
    $C_{risk} \leftarrow P_f \cdot 10,000 \cdot C_r + C_s$
    $r(b(s_t), \mathbf{a}) \leftarrow C_{action} + C_{risk}$
    $P_f, P_s \leftarrow updateRisks(b(s_{t+1}))$
    if t = T **then**
      done is *True*
    end if
    Return tuple $(b(s_{t+1}), r(b(s_t), \mathbf{a}), done)$
    $b(s_t) \leftarrow b(s_{t+1})$
  end for
end for

Algorithm 3 Proximal Policy Optimization training

Initialize critic network: $\theta_v \leftarrow \theta_0$, Initialize actor network: $\theta \leftarrow \theta_0$.
for $k \leftarrow [0...K]$ **do**
  $\mathcal{D}_k \leftarrow$ collect trajectories τ_i from BridgeEnvironment (algorithm 2) given $\pi_k = \pi(\theta_k)$
  Estimate advantages $\hat{A}_t$ using any advantage estimator algorithm based on V_{θ_v}
  Update critic network using MSE, typically via gradient descent algorithm.:

$$\theta_{v,k+1} \leftarrow \arg\max_{\theta_v} \sum_{i \in \mathcal{D}_\parallel} \sum_{t=0}^{T} (v(s_t|\theta) - r(b(s_t), \mathbf{a})) \tag{5.12}$$

  Update actor network using clipped loss function, typically via gradient ascent algorithm:

$$\theta_{k+1} = \arg\max_{\theta} \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^{T} \left(\frac{\pi_\theta(a|s)}{\pi_{\theta_k(a|s)}} \hat{A}_t, \text{clip}(\frac{\pi_\theta(a|s)}{\pi_{\theta_k(a|s)}}) \pi_{\theta_k(a|s)} \hat{A}_t \right) \tag{5.13}$$

end for

5.3.2. Benchmarks

Most of the available methods used for inspection and maintenance planning simplify the problem by working on a component level, assuming independence in costs, structural behavior, and statistics. This simplification is implemented due to the high computational efforts required to find global and optimal planning on a system level.

The traditional methods used for the benchmarks are heuristic approaches. Heuristic approaches are typically used when finding the optimal solution is too complex and an approximation is seen as sufficient. The optimization is done with predefined heuristic rules to limit the total amount of possible policies, however, the optimality of such an approach depends greatly on the designer's experience with such a problem [39]. An example of a heuristic approach is to perform an inspection every five

years.

To verify and evaluate the optimal policy found by the framework, benchmarks are required to set a threshold for the framework to beat. As has been covered in the introduction of section 2.3, the most common approaches are time-based and condition-based benchmarks. The benchmarks are based on heuristic decision rules. Examples of such heuristic decision rules are given below.

1. The time intervals between different inspection and maintenance actions.
2. The age of a component.
3. The damage state of each component.
4. The probability of failure of the structure

In a time-based maintenance benchmark (TBM), a grid of different time intervals is set up. The goal is to find the optimal threshold of time intervals for all possible maintenance actions. An example of a time-based benchmark is, for instance, repairing a component every 5 years and replacing a component every 15 years. Such a system is an example of systematic preventive maintenance, where the condition of the environment is not taken into account in the decision-making [64].

The condition-based benchmark (CBM) consists of a grid of time intervals for the inspections and a grid of damage states for the maintenance actions. The damage states act as a performance metric of the component. The grid of damage states for the maintenance actions is the same as the discretization of the damage state of the case study. The starting state is 0 damage and the final state is 20% damage of the cross-sectional area.

Both for the CBM and TBM benchmarks, the following approach has been applied to find the optimal thresholds for either the time intervals for TBM or time/condition intervals for the CBM. Owing to the stochastic nature of the deteriorating environment, many simulations have to be run for each benchmark. A benchmark for both the TBM and CBM for each climate scenario has been established by performing numerous simulations until convergence was found. Next, the optimal threshold was simulated 10,000 times to establish a confidence interval for the benchmarks.

In table 5.8, the time-based benchmarks are described based on their optimal thresholds, an average lifecycle cost, and a lower- and upper bound for the 95% confidence interval. It can be observed that there is no change in the thresholds for maintenance actions between the climate scenarios. The optimal time-based strategy is to perform two minor repairs and one major repair during the lifetime of a component.

	Interval		*LC costs* $[10^5€]$		
Scenario	**Minor**	**Major**	**Average**	**Lower**	**Upper**
gl	8	16	17.01	16.66	17.8
gh	8	16	17.03	16.67	17.81
wl	8	16	17.05	16.69	17.83
wh	8	16	17.07	16.69	17.9
traditional	8	16	16.96	16.64	17.63
random	8	16	17.02	16.67	17.81

Table 5.8: Time-based benchmarks for the different climate scenarios

In table 5.9, the condition-based benchmarks are described based on their optimal thresholds for inspection- and maintenance actions, average lifecycle costs, and a lower- and upper bound for the 95% confidence interval. It can be observed that the same thresholds are used for each climate scenario. This is in line with the observations made during the deterioration modeling where it was observed that the differences between the scenarios in terms of damage were minor. Another observation is that neither the average LC costs nor the spread of the confidence interval change much between the scenarios. According to the CBM benchmark, an inspection is always followed by a minor repair. Only if the damage is bigger than the major repair threshold, a replacement is performed.

	Interval			*LC costs* $[10^5€]$		
Scenario	**Inspection**	**Minor**	**Major**	**Average**	**Lower**	**Upper**
gl	10	0	0.044	16.38	15.91	16.9
gh	10	0	0.044	16.41	15.93	16.93
wl	10	0	0.044	16.43	15.95	16.93
wh	10	0	0.044	16.45	15.95	16.95
traditional	10	0	0.044	16.32	15.87	16.85
random	10	0	0.044	16.4	15.91	16.93

Table 5.9: Condition-based benchmarks for the different climate scenarios

By comparing TBM with CBM, the following observations can be made. First, CBM always outperforms TBM. This is expected since CBM used performance metrics of the actual state of the structure and can therefore take more risk in the decision-making. This extra risk in decision-making is also reflected in the interval between I&M actions. Since CBM only performs an inspection every 10 years, the total amount of maintenance actions per component during the lifetime is 2 compared to the 2 minor repairs and 1 major repair of the TBM. One might suspect a bigger spread in the confidence interval of the lifecycle costs since fewer maintenance actions are taken, however, this is not the case. The CBM shows less variance due to the value of information of performing an inspection prior to performing a maintenance action.

5.3.3. Results

It has been discussed what algorithm will be applied to find the optimal policy and how to verify the results by means of the benchmarks. In this subsection, the algorithm will be applied to the case study, the training process given and the results presented and discussed. As stated in the introduction of the algorithm, SB3 [62] has been used to set up the PPO model. In table 5.10, the hyper-parameters used in the setup of the model are given. The input for both the actor and critic network consists of 205 input nodes. The first 50 nodes are for probability mass functions over the damage states of the components. The next 150 input nodes are used to describe the deterioration rate of each component. The last 5 nodes are used for the belief over the climate scenarios. The output of the critic network is the value function over the state of the environment and is used to compute the advantage function in the objective function of the algorithm. The output of the actor-network is five probability mass functions over the possible actions per component. The resources used to train the different models were provided by the Delft High-Performance Computing Centre (DHCP) [65]. Each independent initialization was trained on 4 Intel XEON E5-6248R 24C clocked at 3.0GHz and 1GB RAM per CPU and trained for 5 days.

Hyper-parameter	**Value**
Learning-rate	0.0003
Gamma	0.99
Lambda	0.95
Clip range	0.2
# inner layers	2
Size inner layer	128
# environments	4

Table 5.10: Hyper-parameters of PPO

In reinforcement learning, the training is dependent on the policy because the agent is generating its own training data by interacting with the environment. Therefore, the data distributions over the observations and rewards change while the agent is learning. This causes RL algorithms to be sensitive to hyperparameter tuning and initialization. In order to bypass problems with initialization, the chosen framework was trained on five independent realizations. The training process of the five runs is given in figure 5.6. In the figure, the y-axis is the normalized lifecycle costs ($\cdot 10^5$€) and the x-axis gives the number of episodes. The LC costs of each run are averaged based on 100 episodes. Since

the framework was able to beat both the CBM and TBM benchmarks, no further tuning of the hyper-parameters was performed.

Figure 5.6: Five PPO runs applied to case study

From figure 5.6, it can be observed that each initialization is able to beat the TBM benchmark. The CBM benchmark is beaten by all but one run somewhere during the training process.

The goal of the framework is to include the partial observability of climate change within the decision-making of inspection and maintenance planning. In order to draw a conclusion on whether the framework is able to use the different climate scenarios within the decision-making, the best policy from each run is simulated 1,000 times to compare the different policies and look for similarities. The results are summarized in tables 5.11 to 5.15.

Scenario	Inspect	Minor	Major	Average
gl	14.03	4.778	5.267	-16.35
gh	12.20	4.679	5.379	-16.43
wl	12.38	4.658	5.369	-16.39
wh	11.94	4.633	5.411	-16.44
traditional	14.47	4.794	5.275	-16.33

Table 5.11: Policy for run 1

Scenario	Inspect	Minor	Major	Average
gl	11.51	4.105	6.021	-16.36
gh	11.38	4.117	6.024	-16.35
wl	11.97	4.203	6.054	-16.41
wh	11.31	4.164	6.093	-16.45
traditional	11.04	4.13	5.994	-16.32

Table 5.12: Policy for run 2

Scenario	Inspect	Minor	Major	Average
gl	13.08	5.757	5.264	-16.37
gh	14.20	5.729	5.29	-16.40
wl	12.33	5.758	5.264	-16.39
wh	12.32	5.665	5.36	-16.46
traditional	12.24	5.912	5.133	-16.32

Table 5.13: Policy for run 3

Scenario	Inspect	Minor	Major	Average
gl	11.30	6.98	5.037	-16.33
gh	11.67	6.942	5.067	-16.33
wl	10.38	7.933	5.075	-16.30
wh	10.32	7.972	5.036	-16.34
traditional	9.139	7.864	5.132	-16.27

Table 5.14: Policy for run 4

Scenario	Inspect	Minor	Major	Average
gl	11.89	4.092	5.94	-16.47
gh	12.59	4.028	6.01	-16.52
wl	14.94	4.037	6.005	-16.54
wh	14.00	3.988	6.056	-16.58
traditional	15.43	5.061	4.979	-16.39

Table 5.15: Policy for run 5

From the tables, the following observations can be made. The worst policy is found by run 5 in table 5.15. This policy prefers to do more inspections when climate change becomes more prevalent. Especially for scenario wh, the average life cycle cost deviates the most from the benchmarks. The best policy is provided by run 4 in table 5.14. This policy prefers to do inspections for climate change scenarios gl and gh and an increase in maintenance actions for scenarios wl and wh. Run 1 (table 5.11) and run 3 (table 5.13) show similar behavior as run 4. An interesting observation is that the runs which adapt their policy on the amount of climate change seem to outperform the runs which do not change their policy based on the climate scenarios, indicating that an optimal policy does make a distinction between the different climate change scenarios. Another observation that can be made is the similarity between the policy realizations for runs 1 to 3. For run 2 in table 5.12 it appears that the agent assigns a low weight to the belief over the climate scenarios. This can be seen by the similarity in decision-making under the different climate scenarios. One interesting result which is found is the small difference between the traditional scenario and the other climate scenarios for the best policy realization. Since run 4 is able to outperform the other runs and shows the most diverse policy for the different scenarios, this run is selected as the optimal policy and considered for further analysis.

In table 5.16, the optimal policy is compared to the benchmarks. From the table, it can be observed that the optimal policy does not beat the benchmarks by a long shot. An interesting observation is that the optimal policy seems to outperform the benchmarks at an increasing rate for scenarios with more climate change.

It is suspected that the small differences between the benchmarks and the policy are due to the simplicity of the case study. The case study is for instance modeled as a series system, which assumes that if one component fails, the entire bridge fails. The components are also similar in terms of failure modes and degradation. The thresholds of the benchmarks are independent of the components too, which is another factor. Also, as stated multiple times, the differences between the climate scenarios in terms of damage are marginal, which reduces the complexity too. All of these factors play hand-in-hand with the benchmarks which are not able to capture such complexities as a policy found by DRL can.

Scenario	Average	Lower	Upper	Relative to CBM [%]	Relative to TBM [%]
gl	-16.33	-17.22	-15.94	99.72	96.02
gh	-16.35	-17.12	-15.92	99.62	95.99
wl	-16.29	-17.04	-15.88	99.15	95.55
wh	-16.37	-17.39	-15.90	99.48	95.87
trad	-16.24	-16.91	-15.87	99.52	95.76

Table 5.16: Policy compared to benchmarks

A single realization of the optimal policy is given in figure 5.7. From the realization it is evident that the agent appears to show intelligent behavior by, for instance, taking no inspection or maintenance actions at the start of its lifetime and increasing the frequency of inspections nearing the end of the lifetime. It also appears that the agent applies a similar policy for similar components (e.g. components 2 and 4). The agent does not allow large damages to occur on any of the components due to the large penalty it gets on the risk front. Since a single realization is not able to capture the stochasticity involved in the deterioration process, two extra figures have been provided. In figure 5.8, the policy provided by the framework is compared against the benchmarks in terms of actions taken per lifetime, averaging over 1,000 episodes. In figure 5.9, the policy for different components given climate scenario 'wh' is given, averaged over 1,000 episodes.

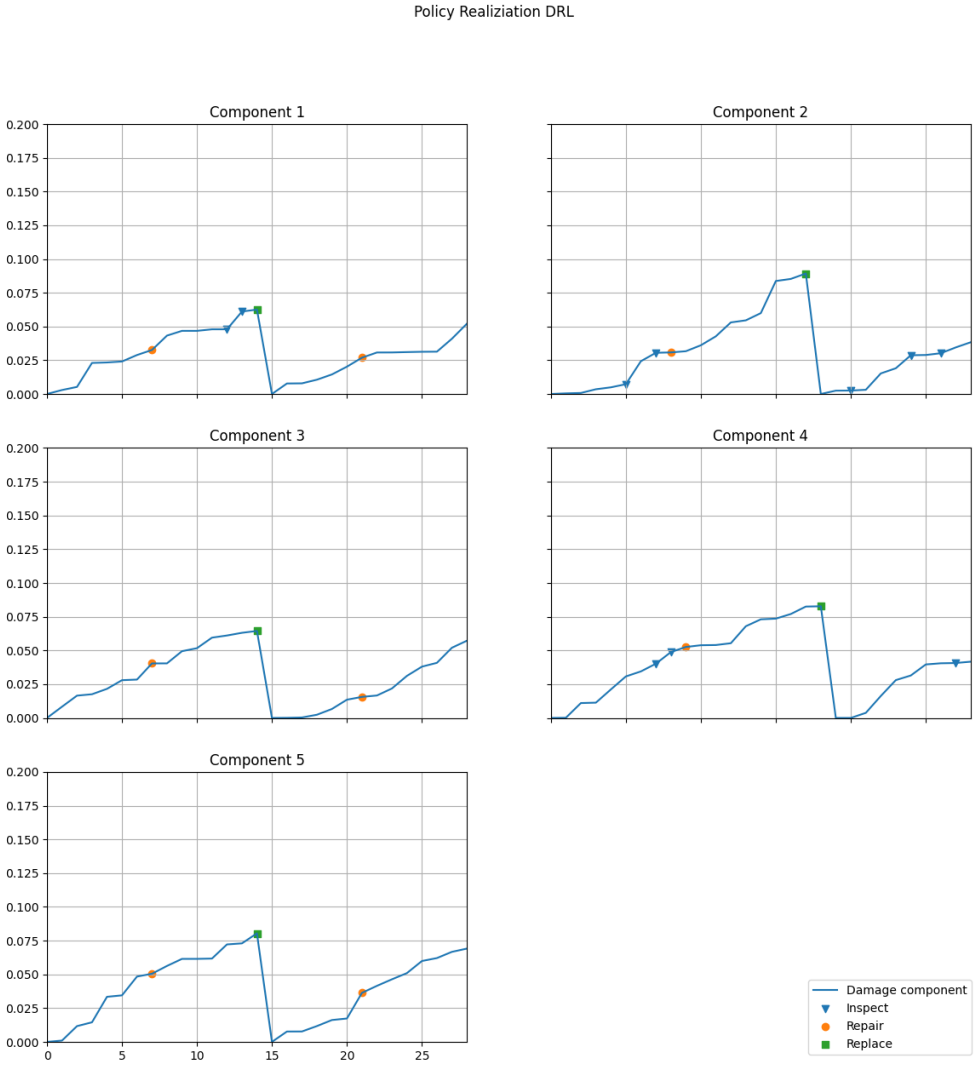


Figure 5.7: Policy realization under framework

In figure 5.8, the actions taken by the optimal policy are compared against both benchmarks for the different climate scenarios. The TBM makes no distinction between the actions taken for each climate scenario. This is because TBM performs preventive maintenance actions and does not consider the actual damage state of each component. CBM shows a slight change in the number of maintenance actions taken by an increase in major repairs compared to the minor repairs for climate scenarios with more temperature change. The optimal policy found by the framework does make a distinction in the policy for a given climate scenario and shows its adaptability to a changing climate. For the intermediate climate scenarios 'gl' and 'gh,' it tries to save costs by performing extra inspections and reducing the number of minor repairs. However, for scenarios 'wl' and 'wh', it prefers to increase the number of minor repairs. Since a minor repair also comes with an inspection, the agent still keeps an updated belief over the states and finds a balance between the actions. Interestingly, the optimal policy tends to perform a similar policy for the worst climate scenario 'wh' and the traditional climate scenario. This could result from multiple local optima being present since the change in damage between the climate scenarios is minimal. This could explain why a similar policy is taken between the traditional and wh scenarios.

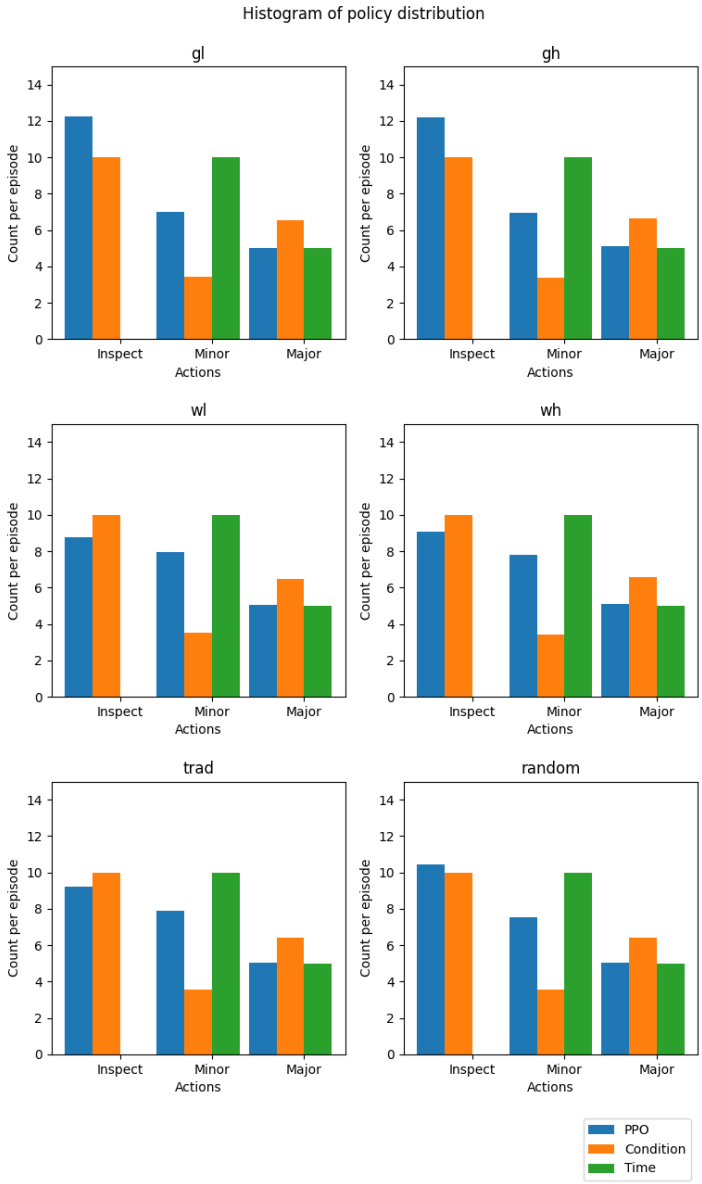


Figure 5.8: Histogram of the actions under different climate scenarios

From figure 5.9, the following observations can be made. The cross-sections at the first span (CC1 and CC5) show different policies while being identical in terms of physical properties. This could be the result of the agent being stuck in a local optimum or that the training process has not yet converged to the optimal policy. The cross-sections at the supports (CC2 and CC4) show a very similar policy with a significant increase in the number of inspections. The cross-section in the second span (CC3) has the least amount of I&M actions performed, which was to be expected since this component has the lowest effect on the probability of failure. Another interesting observation is the tendency of the agent to perform more minor repairs for components CC1, CC3, and CC5 which are situated at the bottom of the slab. Since the thickness of the bottom cover is smaller than that of the top, the probability of spalling is higher. By performing a minor repair, spalling is effectively reset which explains this observation. From the figure, it is clear that the policy tries to find patterns within its decision-making for the different components. For instance, CC2 and CC4 have the highest influence on the probability of failure which results in an increase in the number of inspections and CC3 has the lowest influence which results in the least amount of actions performed. The figure clearly shows that the agent is able to capture the complexity of the different components, which is something that the heuristic benchmarks can not. This added complexity becomes more evident when the components used are not as similar as was present in the case study.

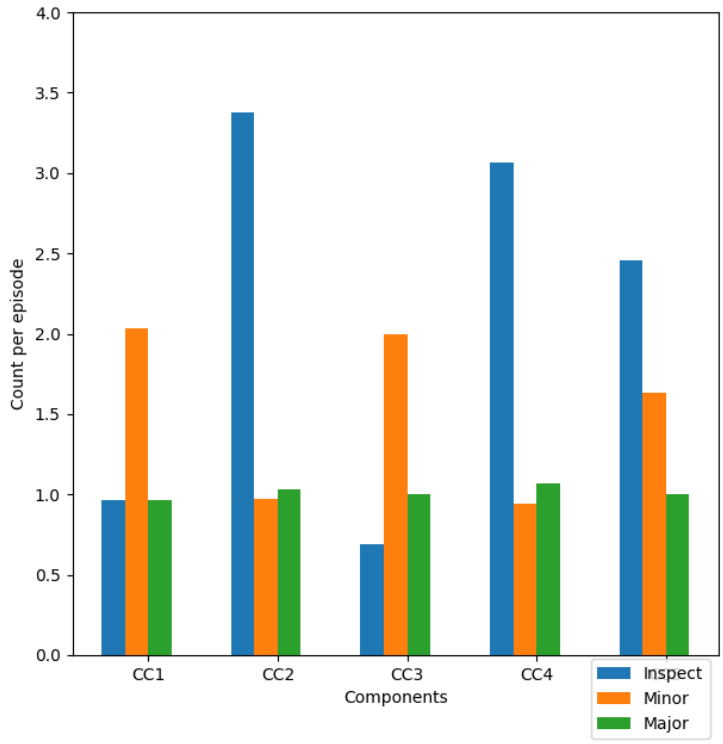


Figure 5.9: Histogram of actions per component under climate scenario wh

From the results, it has become evident that the policy found by the framework is able to capture added complexity compared to the benchmark policies. The first bit of extra complexity is the ability to adapt the policy based on the state of the climate. This can be seen by the agent taking different I&M actions for the different scenarios. The second bit of extra complexity of the framework is the ability to find different policies for the components. This can be seen clearly by the number of maintenance actions for the components at the bottom of the slap compared to the components at the top of the slab.

One assumption taken at the set up of the benchmarks is that it is known what climate scenario the benchmark is subjected to. However, this is not the case in real life. Typically, the benchmarks are set up based on no climate change which is represented by the 'traditional' climate scenario. Since

the intervals of the heuristics for the benchmarks show no change between the different scenarios this conclusion has sadly not become evident. Also, since the components used in this study are relatively similar, the added complexity of the framework to capture such changes has gone unnoticed in terms of life-cycle costs. This is however only the result of the case study being a 'bad' fit for the framework and it is therefore recommended to apply the framework to a case study with more complexity. An attempt has been made to add such complexity to the climate scenarios of the study in the next section which contains an extra analysis.

5.4. Extra analysis

With the results of the framework to the case study known, it has become evident that the framework does not come to its full potential due to the limited effect of the climate scenarios on the damage of the cross-section. In this section an extra analysis will be discussed where the effects of climate change on the damage have been illustratively enhanced since the effects of climate change can occur on a multitude of risk categories [3]. In this analysis, everything apart from the corrosion rate is kept the same as in the previous analysis and therefore only the results will be given and discussed.

In the case study, climate change was present in the setup of the SPD model for corrosion and in the cumulative distribution function for spalling by means of a temperature-dependent corrosion rate, see equation 2.2.

In order to highlight the adaptability of the framework to different climate scenarios, the following approach has been taken. Equation 2.2 adopts a factor K_T which expresses the importance of temperature change to the corrosion rate. This factor has been increased until the difference between the traditional scenario and the wh scenario is about 50%. The equation has also been altered to have the same starting point as the initial analysis of ±0.20 damage by means of a factor K_0.

$$i_{extra}(t) = i_{corr,20} \cdot K_0 \cdot [1 + K_T(T(t) - T_{trad})] \tag{5.14}$$

where: $i_{extra}(t)$ is the corrosion current density used for the extra analysis, K_0 is a factor to initialize the initial damage condition, K_T is a factor to represent the effect of temperature change and T_{trad} is the temperature under no climate change.

The same steps and analysis have been performed as described in the previous three sections. In this section, only the results will be given and additional changes highlighted. The change in corrosion rate causes the following results to be altered

- The stochastic deterioration process 5.1.
- The cumulative distribution function for spalling 5.1.
- The transition models for the POMDP.

5.4.1. Corrosion and spalling

Table 5.17 gives the results used for this extra analysis. The values below correspond to the random variable 'remaining cross-sectional area at the end of the lifespan $A_s(t = 30y)$' for 100,000 samples of the stochastic distributions given in tables 4.2, 4.3 and 4.4. The table presents the average value and a lower- and upper bound corresponding to a 95 % confidence interval.

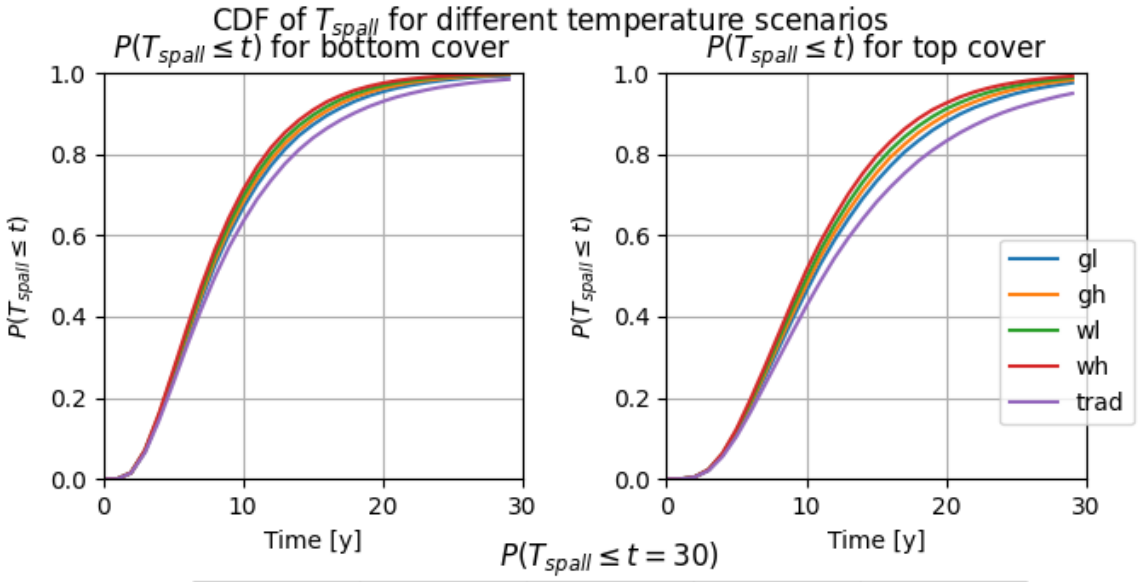
Scenario	average	lower bound	upper bound
gl	238.4	51.73	299.9
gh	231.5	26.06	298.8
wl	224.4	-1.031	297.7
wh	215.9	-33.99	296.3
trad	253.2	106.4	302.4

Table 5.17: Description of the random variable $A_s(t = 30y)$ for the extra analysis

The cumulative distribution function $F_{T_{spall}}$ of the random variable T_{spall} (see equation 2.21) also uses the RVD model. Below the results for the time to spalling for each combination of temperature scenario and concrete cover height are described using the average and 95 % confidence interval.

	Top cover			*Bottom cover*		
Scenario	**Average**	**Lower bound**	**Upper bound**	**Average**	**Lower bound**	**Upper bound**
gl	11.5	3	28	8.46	2	22
gh	11.1	3	27	8.23	2	21
wl	10.7	3	25	8.02	2	20
wh	10.4	3	24	7.79	2	20
trad	12.6	3	34	9.13	2	26

Table 5.18: Description of the random variable T_{spall} for the different temperature scenarios and concrete cover heights



$P(T_{spall} \leq t = 30)$

	gl	gh	wl	wh	trad
Bottom	0.99324	0.9957	0.99726	0.99835	0.98414
Top	0.97555	0.98233	0.98752	0.99188	0.94938

Figure 5.10: Cumulative distribution function for spalling under the temperature scenarios for the bottom concrete cover (left) and top concrete cover (right) for the extra analysis

The gamma process is set up by simulating 10,000 random runs for each initial cross-sectional area and temperature scenario. The initial cross-sectional areas vary between +40% of the starting area ($439mm^2$) and -20% of the starting diameter ($251.3mm^2$) discretized in 30 evenly spaced states. The gamma-process is modeled in terms of damage which is the ratio of damaged cross-section versus initial cross-section (see equation 5.3). In table 5.19, the expected damage with a confidence interval of 95% is given for each temperature scenario under a starting diameter of $20mm$. Figure 5.11 plots the cross-sectional damage versus the time for both the RVD model and the SPD model for each climate scenario.

Scenario	Variable	average	lower bound	upper bound
gl	Damage	0.224	0.141	0.325
	Area [mm2]	243.8	212.1	269.9
gh	Damage	0.244	0.153	0.352
	Area [mm2]	237.5	203.6	266.1
wl	Damage	0.266	0.167	0.386
	Area [mm2]	230.6	192.9	261.7
wh	Damage	0.29	0.182	0.421
	Area [mm2]	223.1	181.9	257.0
trad	Damage	0.179	0.112	0.26
	Area [mm2]	257.9	279.0	232.5

Table 5.19: Description of the variable $d_j(t = 30y)$ for the different temperature scenarios.

From the tables and figures presented in this subsection, the following observations can be made. First, it can be seen that the altercation of the equation of the corrosion rate effectively increases the change between damage under the different climate scenarios. In figure 5.11, this change can clearly be seen by comparing the means. Second, it can be seen that the time to spalling for the climate scenarios has also changed. This change, however, is relatively small compared to the change in cross-sectional damage. It is expected that the change in damage between the scenarios leads to different policies required under the climate scenarios.

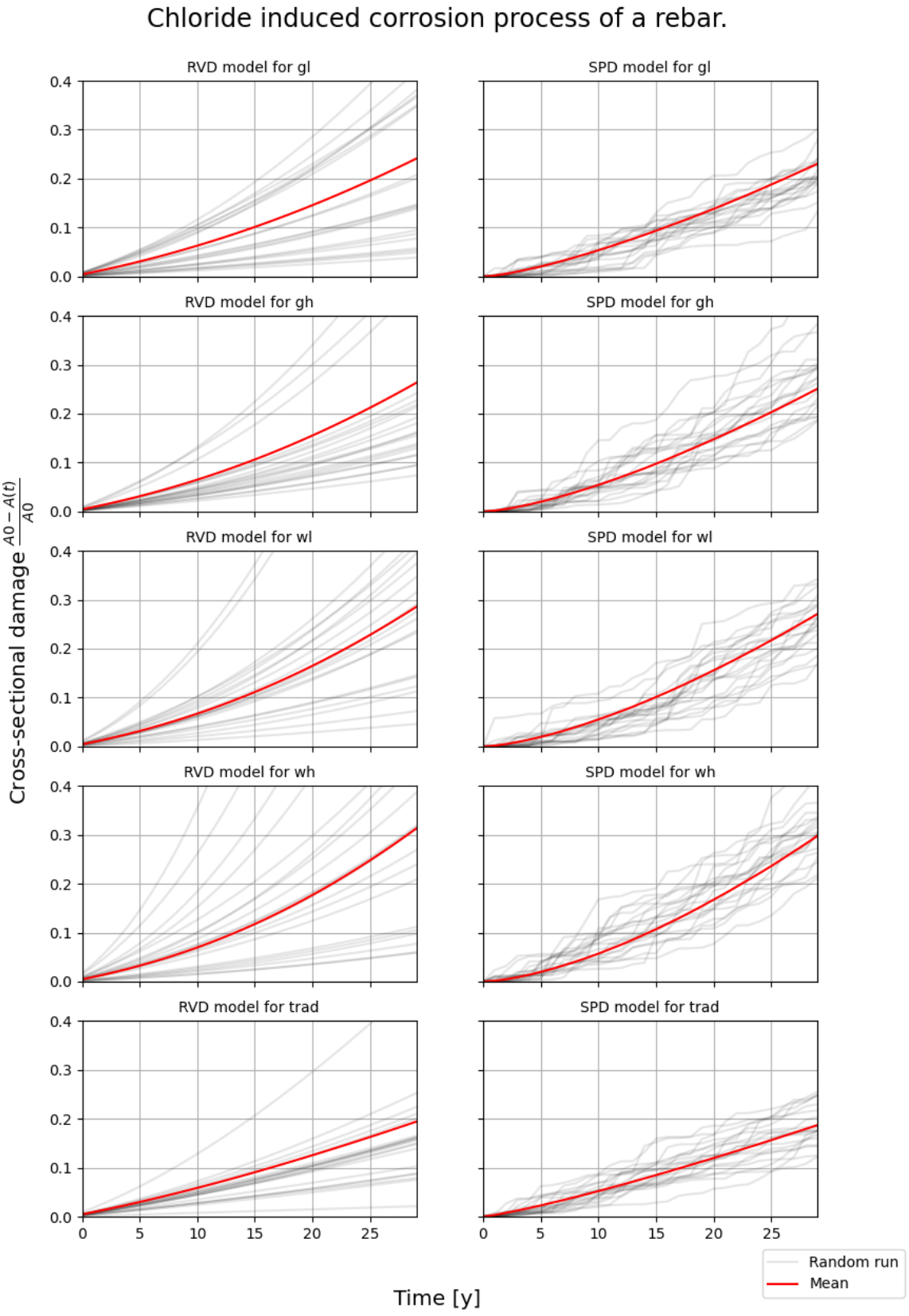


Figure 5.11: Chloride induced corrosion process for Random Variable Degradation model (RVD) (left) and Stochastic-Process Degradation model (SPD) (right) for the different climate scenarios for the extra analysis.

5.4.2. Benchmarks

The same TBM and CBM have been constructed for every climate scenario for the new analysis. Again, benchmarks were constructed. The benchmarks are computed using an exhaustive policy search for each climate scenario. Convergence of the values is ensured by implementing stop criteria over the mean value of the cheapest benchmark. Once the optimal heuristics have been found, each optimal threshold has been simulated 10,000 times to establish a confidence interval for the benchmarks.

	Interval		*LC costs* $[10^5€]$		
Scenario	**Minor**	**Major**	**Average**	**Lower**	**Upper**
gl	-	10	16.71	16.68	16.74
gh	-	10	16.82	16.75	16.86
wl	-	10	16.99	16.93	17.04
wh	-	10	17.08	17.02	17.13
traditional	10	20	16.28	15.39	18.38
random	-	10	16.8	16.39	17.11

Table 5.20: Time-based benchmarks for the different climate scenarios

	Interval			*LC costs* $[10^5€]$		
Scenario	**Inspection**	**Minor**	**Major**	**Average**	**Lower**	**Upper**
gl	10	0	0.044	16.29	15.78	16.81
gh	10	0	0.044	16.5	15.98	16.94
wl	10	0	0.044	16.78	16.3	17.13
wh	10	0	0.044	16.95	16.47	17.24
traditional	10	0	0.044	15.69	15.24	16.24
random	10	0	0.044	16.44	15.39	17.2

Table 5.21: Condition-based benchmarks for the different climate scenarios

Table 5.20 shows the optimal intervals for the maintenance actions for the different climate scenarios. It also gives the LC costs in terms of an average and a lower- and upper bound for the 95% confidence interval. The first observation that can be made is that a different policy is taken for the scenarios with climate change versus the ones without climate change. The one without climate change performs one minor repair and one major repair during its lifetime. It is interesting to note that this policy is different than the one in the prior analysis. The saving in terms of actions performed does come at a cost using a significant increase in the variance of the LC costs. The policy for scenarios under no climate change performs two major repairs during its lifetime, which comes with a small variance. This small variance results from the major repair resetting the damage state and deterioration rate of the given component.

Table 5.21 shows the optimal thresholds in terms of years for the inspection and in terms of damage for the actions. The first clear observation is that the same intervals in the heuristics are used for the different scenarios. Therefore the extra analysis cannot capture the complexity of climate scenarios required for the framework to work optimally. Nevertheless, the increase in damages should lead to different policies being applied to the scenarios with the worst climate change scenarios preferring major repairs over minor repairs.

5.4.3. Results

The training for extra analysis shows the tendency of DRL algorithms to be dependent on initialization. From figure 5.12, a much more 'chaotic' training process can be observed. Run 2 and run 5 have instabilities in the training process and appear to adopt an inefficient policy which can be seen by an instant increase in the average LC costs. Also, the policies appear not to beat the CBM benchmarks this time, except for run 1.

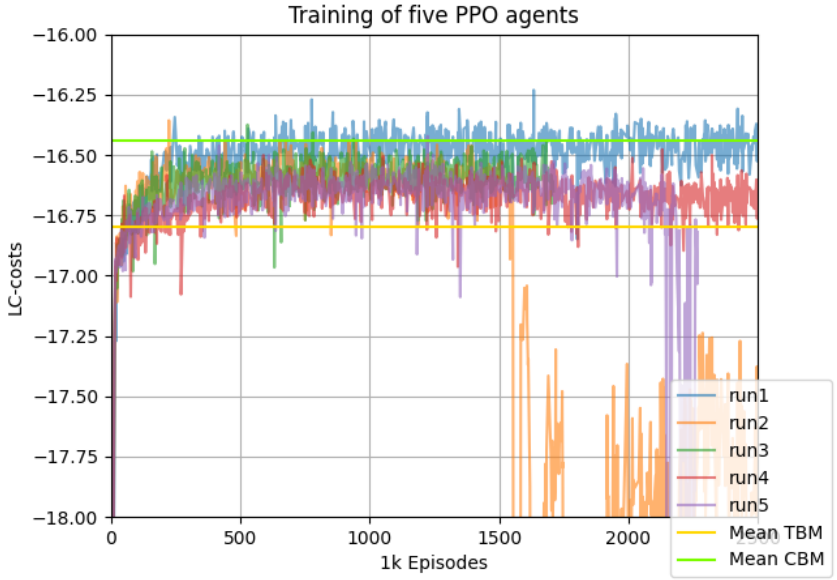


Figure 5.12: Five PPO runs for extra analysis

In tables 5.22 to 5.26, the policy for each of the runs is given. The first observation that can be made is the wide range of adopted policies for the different runs. The runs can be grouped into two categories, especially ones that prefer inspections over maintenance actions (runs 1 and 2) and those that prefer performing maintenance actions (runs 3 to 5). The runs which prefer inspections over maintenance actions also tend to perform more minor repairs, which also come with an inspection. These runs are, therefore, close to the CBM benchmarks, which adopt a condition-based approach and runs 3 to 5 are closer to the TBM approach, which prefers preventive major repairs. The run with the lowest LC costs is run 1 and is adopted as the optimal policy for further analysis.

When comparing the policies from the runs of the extra analysis (tables 5.22 to 5.26) with the prior analysis (tables 5.11 to 5.15) the following observations can be made. The first thing is the homogeneity of the policies adopted by the first analysis. The policies of the prior analysis show a preference to perform inspections versus maintenance actions, with the most significant difference being the split between minor or major repairs. The second observation is the big difference in the policy for the traditional runs, with the optimal policy of the last analysis largely performing inspections over maintenance actions. When comparing the optimal policy of both analyses (run 4 in table 5.14 for the initial analysis and run 1 in table 5.22 for the latter), it becomes evident that an increase in climate change leads to a rise in the number of maintenance actions and a decrease in the number of inspections.

Scenario	Inspect	Minor	Major	Average
gl	15.23	3.47	6.59	-16.30
gh	11.62	3.11	6.95	-16.47
wl	8.51	2.78	7.29	-16.74
wh	8.17	2.17	7.94	-17.02
traditional	20.23	4.69	5.35	-15.83

Table 5.22: Policy for run 1 (extra analysis)

Scenario	Inspect	Minor	Major	Average
gl	14.00	3.12	6.90	-16.33
gh	13.24	2.85	7.20	-16.55
wl	8.67	2.33	7.72	-16.79
wh	9.28	1.91	8.21	-17.06
traditional	14.22	2.64	7.41	-16.12

Table 5.23: Policy for run 2 (extra analysis)

Scenario	Inspect	Minor	Major	Average
gl	6.97	1.79	8.23	-16.46
gh	6.22	1.67	8.35	-16.61
wl	4.82	1.50	8.53	-16.85
wh	5.40	1.12	8.93	-17.00
traditional	18.63	3.89	6.16	-15.98

Table 5.24: Policy for run 3 (extra analysis)

Scenario	Inspect	Minor	Major	Average
gl	6.72	0.83	9.17	-16.60
gh	7.15	0.71	9.31	-16.75
wl	4.46	0.69	9.46	-16.98
wh	4.43	0.72	9.57	-17.22
traditional	14.38	3.90	6.11	-15.87

Table 5.25: Policy for run 4 (extra analysis)

Scenario	Inspect	Minor	Major	Average
gl	7.17	1.90	8.15	-16.42
gh	6.14	1.76	8.28	-16.58
wl	5.45	1.71	8.34	-16.80
wh	5.19	1.40	8.66	-16.98
traditional	19.85	2.00	6.03	-16.31

Table 5.26: Policy for run 5 (extra analysis)

In figure 5.13, a single policy realization is given for the traditional climate scenario. During the first ± 10 years, no inspection or maintenance actions are performed and followed by the first round of maintenance actions, with mainly major repairs occurring. Note this is the same policy that the CBM would apply, except for major repairs occurring for components 1 and 2, where CBM would use a minor repair. During the following ten years, many inspections occurred to determine the following action to be taken. If component one is examined, a sub-optimal policy is found. Many inspections occur right after a replacement which resets the component's state, which is illogical. After the second round of maintenance repairs, the policy takes no more actions except for component three, where inspections are performed.

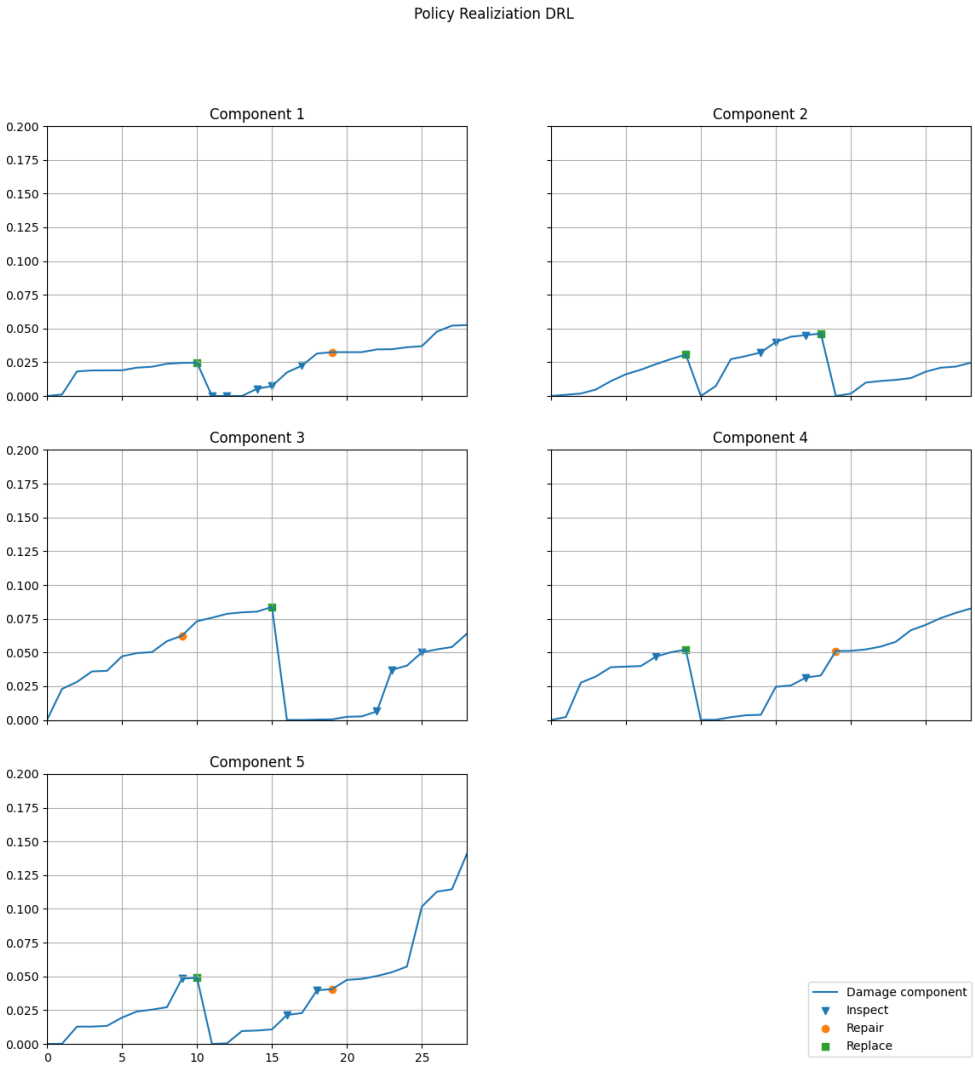


Figure 5.13: Policy realization under traditional scenario

In figure 5.14, a single policy realization is provided for climate scenario wh. The figure clearly shows why the runs cannot beat the benchmarks by a long shot. For scenario wh, the TBM and CBM benchmark are practically the same: to perform a major repair every ten years. CBM and the policy provided by the framework can beat the TBM benchmark due to occasionally performing a minor repair if the structure's condition has only marginally decreased. Major repairs, practically replacements, play hand-in-hand with a preventive maintenance policy that ultimately resets the deterioration process.

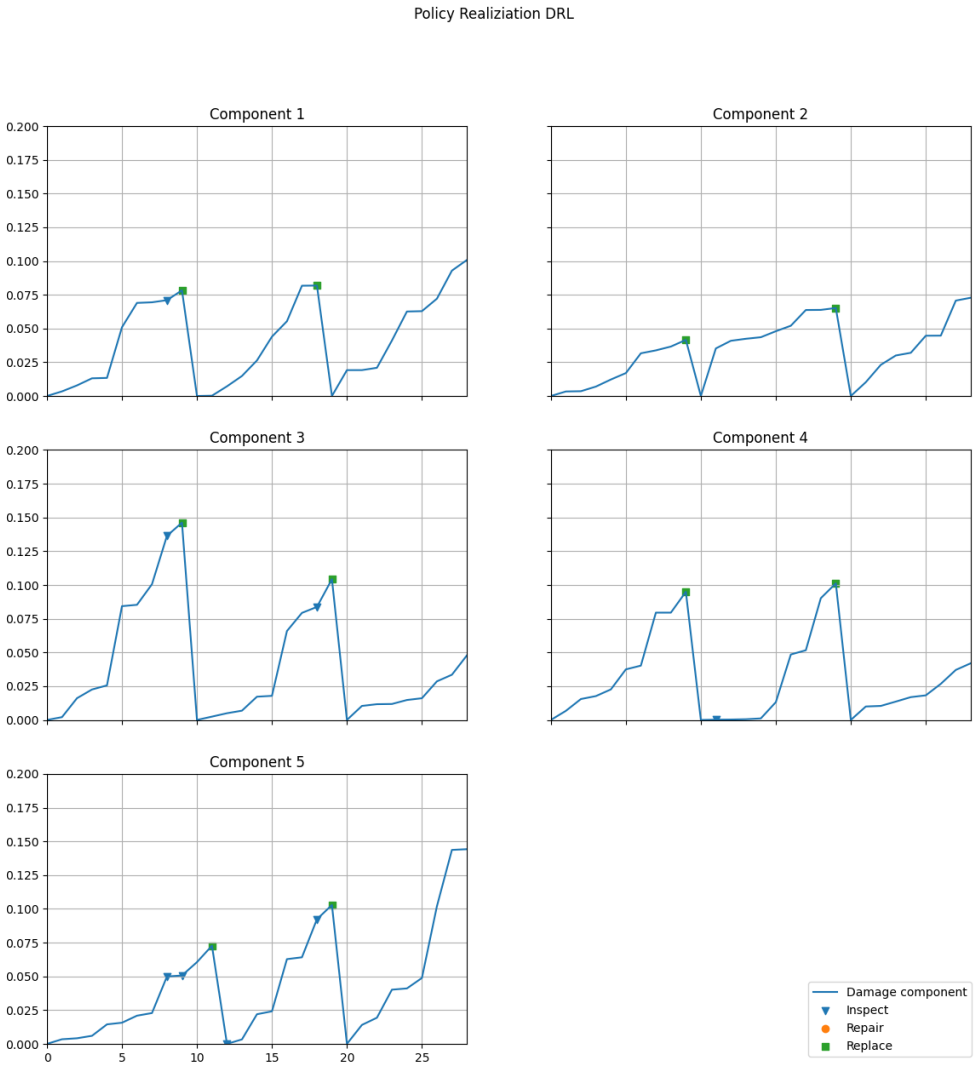


Figure 5.14: Policy realization under wh scenario

Figure 5.15 shows the I&M actions taken by the policy from the framework and the benchmarks for each climate scenario. The first observation to be made is the similarity between the benchmarks. In all climate scenarios apart from the traditional, major repairs far outweigh the number of minor repairs taken. The second observation is the similarity of PPO to both benchmarks regarding the division of actions taken. The histograms show that PPO can adapt its policy based on the belief over the climate scenarios it gets by adapting its policy to the benchmarks.

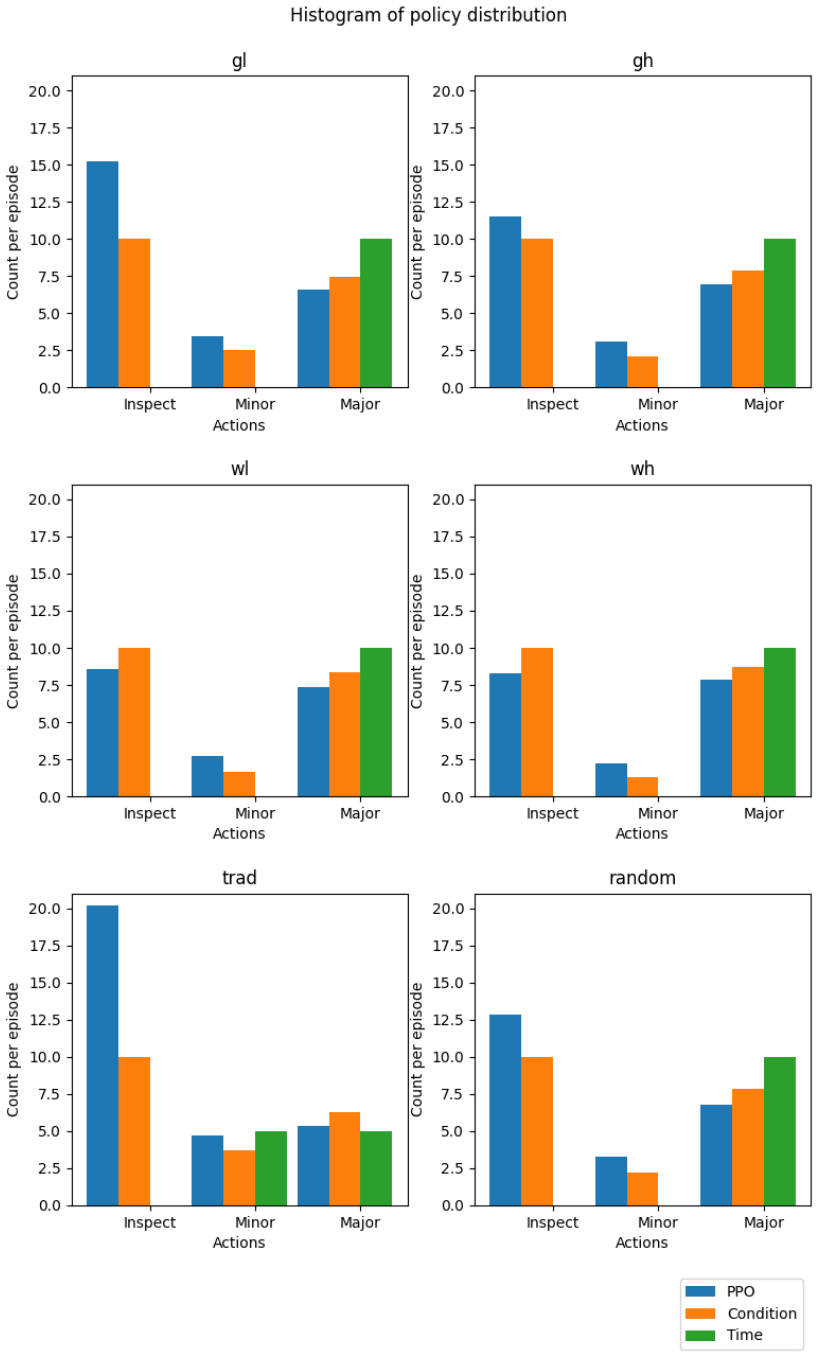


Figure 5.15: Histogram of the actions under different climate scenarios (extra analysis)

5.5. Validation of the framework

In this section, the found policy is validated against the user's wishes set in the goal for the initial analysis of the case study. The goal of the user can be dissected in the following requirements:

1. The failure probability must not exceed 1 % for each decision step.
2. Costs are minimized and lower than the benchmarks.
3. Climate change is considered within the decision-making.

From the case study and the extra analysis, it can be concluded that the framework effectively takes the partial observability of climate change into account in the decision-making of the framework. An example is the increase in major repairs for more extreme climate scenarios for the extra analysis. The framework can also beat both the CBM and TBM benchmark for the initial- and extra analysis despite the ill-fit of the case study to the framework.

First off, the initial assumption of the client to remain below 1% failure probability was not the right measure to compare the policies and validate the framework. The approach is to collect 1,000 random simulations of both the CBM benchmark and the policy of PPO for climate scenario wh to compare the probability of spalling and the probability of failure.

Figure 5.16 plots the failure probability against the planning horizon for both PPO and CBM. The first observation that can be made is that the mean of both policies has the same maximum failure probability. Another observation is the steep drops in failure probability for CBM due to the maintenance actions performed every ten years after an inspection. The failure probability of PPO shows a more spread-out failure probability over the first twenty years of its planning horizon. The cause of this is due to PPO taking different policies for each component, as can be seen in figures 5.7 and 5.9, causing maintenance actions to be performed at different time steps.

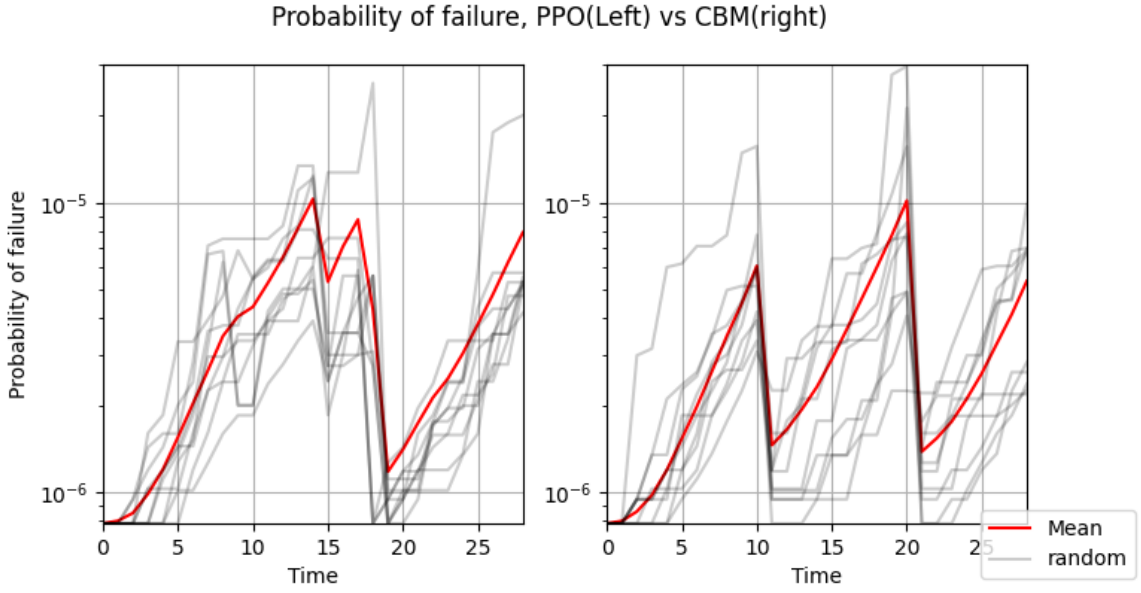


Figure 5.16: Failure probability of PPO vs CBM benchmark

Figure 5.17 plots the spalling probability against the planning horizon for both PPO and CBM. The first observation is that the behavior appears deterministic since no random runs can be identified. This can be explained by the probability of spalling being dependent on the deterioration rate, which is assumed to be known at every time step and therefore is deterministic. The following observation is the probability of spalling approaching 100% for both policies. This observation can be explained by a poor setup of the reward function, with spalling not being penalized enough. However, this is an easy fix by implementing a hard constraint within the POMDP environment to force a minor repair to

be performed if spalling has occurred or penalize the system harder by increasing the cost of spalling. In general, it can be said that the policy found by the framework complies with the user's wishes.

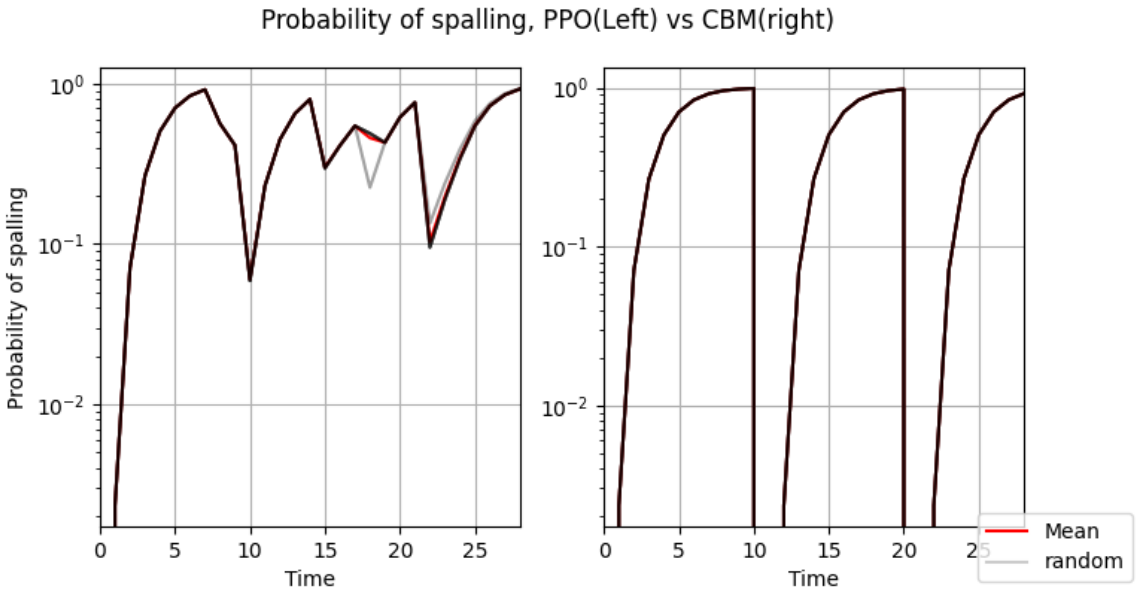


Figure 5.17: Spalling probability of PPO vs CBM benchmark

6

Discussion, conclusion, and recommendations

This thesis has established a framework to incorporate climate change within the inspection and maintenance planning for civil infrastructures. The framework spans three main topics: climate change, stochastic deterioration, and deep reinforcement learning. This chapter provides a reflection on the research process, and recommendations are made. First, in section 6.1, the research problem is concisely introduced, and the main findings are stated. Second, the applied methodology will be discussed in section 6.2. This includes the different phases of the framework, the setup of the case study, and the analysis performed. Third, in section 6.3, the main results gathered from the case study will be presented, and implications introduced. The section compares the traditional approaches and the ones provided by the framework. The results are used to answer the main research question. Fourth, in section 6.4, the framework's implications on policy making for inspection and maintenance planning will be discussed. Fifth, in section 6.5, the study's limitations will be discussed based on the physical modeling, the effect of climate change, and deep reinforcement learning. Sixth, in section 6.6, recommendations for future work in this field will be discussed. The future works can be divided into two categories, case studies, and deep reinforcement learning. And finally, in section 6.7, the research will be concluded by answering the main research questions and giving a final concluding statement.

6.1. The problem

Climate change affects the state of our infrastructure in the following seven risk categories: durability, serviceability, geotechnical, increased demand, accidental loads, extreme natural events, and operational risks [3]. Planning for such risks is difficult due to the many uncertainties and interdependencies related to climate change and the deterioration of our infrastructures.

The first category of uncertainties is regarding the possible future of climate change. To plan for many different futures of humankind, climate scenarios are set up based on greenhouse gas emissions, air pollutant emissions, and land use by the IPCC. The scenarios depend on socioeconomic projections of humankind, for instance, economic and population growth, changes in energy- and land use, and climate policies [12]. However, the climate's actual state will be somewhere between the scenarios.

The next category of uncertainty is to establish the effect of climate change on the structure. The current design of infrastructure does not consider climate change in the design. Therefore, many existing concrete infrastructures are likely to experience a decrease in structural integrity due to climate change. It is therefore recommended to increase monitoring and maintenance for existing concrete structures [6]. The impact of climate change on the infrastructure depends greatly on each structure due to differences in design, local environment exposure, and uncertainties in construction and maintenance, making individual inspection and maintenance planning a requirement [7].

The problem is, therefore, two-fold. On the one hand, there is the problem of identifying the current trajectory of climate change to plan for. And on the other hand, it is difficult to assess the effect of climate

change on the infrastructure in terms of safety, serviceability, and costs. In this thesis, an effort has been made to establish a framework to incorporate the effect of climate change within the inspection and maintenance planning of our infrastructure. The framework includes a belief in the climate, being a probability distribution over the possible climate scenarios, in its decision-making to operate and maintain the infrastructure. An optimal policy in terms of lifecycle costs is found by Deep Reinforcement Learning, which finds efficient lifecycle policies for partially observable large-engineering structures by exploring the state space of the structure. An application of the framework is provided in terms of a case study, where the framework can beat two well-established traditional heuristic-based approaches, namely time-based maintenance (TBM) and condition-based maintenance (CBM), in terms of lifecycle costs. Beating TBM with 5% and CBM with 1% for the respective climate scenarios. The framework has also been shown to alter its decision-making based on the state of the climate by taking more preventive measures in terms of maintenance for scenarios of higher climate change and adopting a more informative approach by increasing inspections for scenarios of lower climate change.

6.2. Methodology

The framework makes use of three phases in the modeling. The first phase translates a physical model (i.e., an engineering structure) towards a goal, boundary conditions, initial conditions, and data collection. The sufficiently described physical model is then translated towards a Partially Observable Markov Decision Process, which acts as the mathematical framework for the sequential decision-making of a structure over its lifetime. To find an optimal policy, which is a sequence of inspection and maintenance actions, one needs to explore an extensive combination of damage states, inspection and maintenance actions, and in this case, climate scenarios. Finding the optimal policy within all possible combinations is similar to finding a needle in a haystack. This problem is known as the curse of dimensionality. The third phase of the framework, finding the optimal policy, applies a Deep Reinforcement Learning algorithm that can efficiently tackle this problem. The algorithm explores the environment through interaction and finds an optimal policy without prior knowledge.

An adaptation of the framework is provided using a case study. The case study is a three-span reinforced concrete slab bridge exposed to chloride-induced corrosion. The effects of climate change are incorporated using a temperature-dependent corrosion rate. The temperature-dependent corrosion rate is used in two limit states, failure of the structure and concrete spalling for a component. The limit states are assessed using a Monte-Carlo analysis using data gathered from existing literature, given in tables 4.1 to 4.6. Climate change is simulated on four climate scenarios for the Netherlands, and one control scenario being no climate change. The order from most to least amount of climate change is as follows: wh, wl, gh, gl, traditional. An optimal sequence of inspection and maintenance actions for the five components is sought in terms of LC costs and consists of maintenance and risk costs. The maintenance costs are found, adding the costs of the actions performed: inspections, minor and major repairs. The risk costs consist of monetization of the probability of spalling and the probability of failure to penalize the system for failure. The policy is compared and verified using traditional heuristic-based approaches, TBM and CBM. The finding of the optimal policy, or in other words, training of the algorithm, is performed using a high-performance computer provided by the Technical University of Delft [65].

The case study provides a good playground for the framework to be applied upon. However, due to the limited scope of the research, not much time could be allocated to modeling a complex engineering structure. A result of this can be seen in minor changes in damage between the different climate scenarios. Especially the initial risk analysis to be conducted in the first phase of the framework is recommended to be extended since this is the basis of establishing the interdependencies on the risks due to climate change. An example for the current case study would be the following; an increase in temperature leads to an accelerated corrosion rate but might also lead to a decrease in the number of frost days. Therefore the exposure to deicing salts decreases, which is the primary catalyst for chloride-induced corrosion. Thus, the framework's potential impact depends greatly on the extent of climate modeling performed a priori. More information regarding the limitations and recommendations for this study will be provided in later sections.

6.3. The results

The results of this thesis have been found by applying the established framework to the case study introduced in chapter 4. In chapter 5, the sub-results are discussed extensively. Therefore this section will primarily focus on the main results and corresponding implications. This section attempts to generalize the results and link them with the main research question.

The first phase of the framework (section 5.1) covered the translation of an engineering structure towards a physical model. The phase covers the setup of climate scenarios, a stochastic deterioration process, and a definition of the actions. It is found that the method to include climate change in the modeling only marginally affects the deterioration of the structure, both in terms of damage and the probability of spalling. This can be seen in table 5.2 and figure 5.3 for the damage of the cross-section, where the difference in terms of damage between no climate change (traditional) and the worst scenario (wh) is a marginal 1.1% of the original cross-section, with 19.6% damage for the traditional scenario and 20.7% for the 'wh' scenario. The effect on the probability of spalling can best be described by the average time required for spalling to occur, represented in table 5.4. The average time for spalling is 14 years for the top cover and ten years for the bottom cover. The change in average time to spalling for the traditional and 'wh' scenario is 0.40 years for the top cover and 0.26 years for the bottom cover. An intelligent policy should implement more maintenance actions for components situated at the bottom cover than the top cover due to a substantial difference in average time to spalling. With a time span of 30 years, spalling would occur twice for the top cover and three times for the bottom cover. The marginal changes between both limit states for the different climate scenarios can lead to the framework not taking any changes in inspection and maintenance actions between different climate scenarios.

In the second phase of the framework (section 5.2), the physical model is described as a POMDP. The failure probabilities presented in table 5.7 show that components CC1, CC3, and CC5 have approximately the same failure probability for each damage state, and components CC2 and CC4 located above the support have double the failure probability. Therefore, an intelligent policy is expected to implement increased maintenance actions for the components near the supports due to the higher influence on the failure probability.

Two traditional heuristic-based approaches have been set up for the third phase of the framework (section 5.3), and an optimal policy for the POMDP is sought for using Deep Reinforcement Learning. The algorithm applied for this task is Deep Centralized Multi-agent Actor-Critic which decentralizes the policy by assuming independence between component actions, significantly reducing the policy space [8]. From the previous phases, it is expected that an optimal policy distinguishes the I&M actions for the different components. The first expectation is an increase in maintenance actions for the components situated at the bottom cover (CC1, CC3, and CC5), especially minor repairs are expected since they effectively reset the spalling process for the lowest price. The second expectation is an increase in maintenance actions for the components above the supports (CC2 and CC4), due to the higher probability of failure, with more major repairs expected due to a complete reset of the damaged state. An optimal policy would therefore perform more minor repairs for CC1, CC3, and CC5 and more major repairs for CC2 and CC4, with identical policies for the first span (CC1 and CC5) and the supports (CC2 and CC4). The third distinction, being the most important for this thesis, is a change in I&M actions for the different climate scenarios. An optimal policy is expected to perform more maintenance actions for scenarios with worse climate change than the traditional scenario.

Due to the minor changes in terms of damage and average time to spalling highlighted in the previous paragraphs, an extra analysis was performed in section 5.4, which illustratively enhances the effect of temperature on the corrosion rate. The effect on the damage is given in table 5.19, with the traditional scenario having 17.9% damage and the wh scenario having damage of 29.0%. The effect on spalling is given in table 5.18. For the top cover, an average time to spalling of 12.6 years has been found for the traditional scenario and 10.4 years for the wh scenario, respectively. The bottom cover has an average time to spalling of 9.13 years for the traditional scenario and 7.79 years, respectively. This analysis has been performed to analyze if the policy can capture the added complexity of the different climate scenarios.

To compare a policy found by the framework against a benchmark. The benchmarks are TBM and CBM, with the former implementing decision rules on time intervals between maintenance actions and

the latter implementing decision rules on maintenance actions based on the condition of a component given an inspection. The benchmarks are compared with the framework regarding decision-making, risks, and LC costs. A comparison of the policies for each climate scenario is provided in figure 5.8, and for the extra analysis in figure 5.15. TBM could not make any distinction in decision-making for the different scenarios or the different components. CBM captured a marginal difference in decision-making for the components due to the scenarios with more climate change surpassing the major repair threshold more frequently.

The policy found by the framework corresponds to run 4 in table 5.14 and run 1 in table 5.22 for the extra analysis, respectively. From run 4, the following conclusions can be drawn. Firstly, from figure 5.9, it can be confirmed that the policy makes a distinction in decision-making between the different components. Confirming the suspicion that an intelligent policy would perform more minor repairs for components at the bottom of the slab and increase the major maintenance for the components located at the supports. Regarding the climate scenarios, the second distinction in decision-making can best be confirmed by figure 5.15. The agent shifts from a condition-based approach for scenarios 'traditional' and 'gl' towards a more preventive maintenance approach for scenarios 'wl' and 'wh.' This can be observed by a gradual transition towards performing more maintenance actions over inspections when sorting the scenarios from most climate change to least. Regarding LC costs, the framework can beat both regular and extra analysis benchmarks. A comparison is provided for the extra analysis in table 5.16, beating CBM to 0.85% and TBM up to 4.5 %. Regarding risk, figure 5.16 and figure 5.17 show the same risk thresholds found for CBM and the framework, proving that the framework does not 'cheap out' on safety.

The results from the previous paragraph highlight the intelligence and adaptability of the framework to capture the different layers of complexity in the case study. This shows that the thesis has accomplished the goal of establishing a framework incorporating climate change in engineering structures' decision-making. One important fact must be highlighted, however. Where the framework can find an optimal policy for the different scenarios or interpolations of them, the benchmarks cannot. The benchmarks are optimized for a scenario; only the traditional scenario is examined in the current practice. This will become more prevalent with a case study where the limit state is more dependent on the climate, leading to more diverse policies for the different climate scenarios. This discussion will be continued in the limitations and recommendations sections. The implications of the framework will be discussed in the next section.

6.4. Implications

The optimality of a traditional I&M policy based on heuristic decision rules depends on the decision-makers knowledge and experience [39]. I&M policies derived from DRL algorithms have been found to converge toward an optimal policy with no prior experience required ([8], [45], [39]). The results from this thesis also support this finding. The policies derived from DRL can capture the different complexities present in an engineering structure and plan optimally for them, as shown by finding lower LC costs than for the benchmark policies. It is therefore expected I&M plannings derived from DRL algorithms can have a significant impact when knowledge and experience in planning for a scenario are limited, for instance, with the impacts of climate change.

In Wang et al. ([5], [6], [7]), it is recommended to increase the maintenance and monitoring of existing reinforced concrete structures due to the effects of climate change. It is, however, also stated that individual I&M plans must be set up due to the varying risks due to location, environmental exposure, and material design. The advantage of DRL algorithms is finding an optimal solution to completely different planning problems with minor to no changes in the network architecture [37]. This finding is also supported by this study, where the same solving algorithm is applied to the extra analysis, and still, an optimal policy is found. This study contributes to a solution for this problem by making it possible to establish policies for individual structures with minor changes to the solving architecture.

In practice, the framework is expected to be used in combination with an engineer, not replacing them. DRL algorithms can change dynamically to the environment or decision-making and act optimally after that. An example is the following, from the framework's policy, it follows that component *x* has to be repaired; the engineer, however, decides from experience that it is better to repair all similar components

in one go since it limits the downtime of the structure. The algorithm can act optimally from thereon and can even quantify the cost of the engineer's action by means of the value function. Therefore, the algorithm is robust to changes implemented by the engineer and can also be used as an informative tool to make well-reasoned decisions.

The ability of the framework to work with model uncertainties is where the real advantage lies with traditional methods. Whereas conventional methods find a policy by all possible combinations of heuristic decision rules for a given scenario, the framework applies a belief over possible scenarios. It can act optimally, also for intermittent scenarios. This gives rise to many possible practical implications, with the study covering the planning for potential climate change futures. Other examples include planning for different loading scenarios, different use cases, deterioration mechanisms et cetera.

Generally, the framework is expected to contribute to the current body of knowledge by providing a foundation for similar studies to be built on. Specifically, making it easier to perform studies on real-life engineering structures to identify the effects on I&M planning. The scope of the thesis was centered around establishing a framework to make planning under uncertainty in (climate) scenarios possible. The resulting framework provides clearly defined steps to be completed to apply such an approach to an engineering structure. The recommendations section will discuss more about the possible future works of the framework.

6.5. Limitations

The framework covers climate change, stochastic deterioration processes, and optimal inspection and maintenance planning. To comply with time restrictions related to the thesis, a scope needed to be defined, leading to some subjects being either simplified or excluded from the study. This section covers the study's primary limitations, with most originating from the simplified modeling in the case study.

The first topic of discussion is the physical modeling of the case study. The physical modeling requires a risk analysis to be performed a priori. The goal of the risk analysis is to establish what governs the failure probability and how they affect the structure's safety. Due to the illustrative nature and limited scope of the thesis, assumptions were taken to have initial- and boundary conditions for the study to be built on. While the assumptions do not affect the ability of the framework to plan for uncertain scenarios, they do affect the comparison with traditional methods. An example can be found in the case study, where the temperature scenarios simulate the effect of climate change on the structure. First, the difference in terms of corrosion between the temperature scenarios is only marginal. Second, it is assumed that only temperature affects the corrosion rate, while many other variables, such as exposure to deicing salts, affect the corrosion rate too. Another example is the assumption of statistical independence between the probability of failure between the components. Risk analysis is expected to contribute majorly to the framework's possible advantages compared to traditional methods and might even be the deciding factor in applying such an algorithm. Therefore, future studies that adopt the framework should allocate a significant portion of time to the risk analysis to be performed a priori.

The second topic of discussion is the conclusions drawn on the effect of climate change. Due to the many interdependencies of the state of the climate and the effect on our infrastructure, it is hard to make conclusions about the exact impact. The results section states that worse climate scenarios lead to more maintenance actions being performed. Although this is in line with existing literature (e.g. [6]), [16]), making firm conclusions is challenging due to only one case study being examined. To make firm conclusions on the effect of climate change on inspection and maintenance planning, it is recommended to perform a comparative analysis of multiple existing infrastructures using the framework proposed in this study.

The third topic of discussion is regarding the DRL part of the work. First, there is the topic of possible DRL algorithms. In this study, only one DRL algorithm has been applied to the case study. It could well be that a different algorithm would have found a more optimal policy, therefore affecting the ability of the framework to beat traditional methods. It is, therefore, recommended to include multiple algorithms in the future if the framework is applied. Second, the computational effort is required to train the agent to find an optimal policy. The case study used to make conclusions about the framework uses a mere five components to limit the computation effort required, whereas complex engineering structures have

components in the hundreds or even thousands. The problem's computational complexity increases exponentially with the number of components, which might lead to the optimal policy being intractable for large engineering structures. However, by introducing assumptions such as independence of actions between components, such as in [8], the total amount of combinations reduces greatly. Especially if this is combined with the recent advancements in DRL and the ever-increasing computational power available, such problems might become history sooner than one can imagine.

6.6. Recommendations

The thesis has established a framework that spans the topics of climate change, stochastic deterioration, and deep reinforcement learning. Since the combination of the three topics is a field not yet explored (up to my knowledge), several recommendations for future work can be made. The recommendations are split into two parts, the case study and deep reinforcement learning.

6.6.1. Case study

The focus of the thesis was on the development of a framework to include climate change within the decision-making of engineering structures. To fully grasp the effectiveness of the framework, it is recommended that the framework be applied to a more elaborate case study. The case study can be made more extensive on the fronts of actions, limit states, or physical modeling. A more detailed case study can provide a better comparison with traditional approaches and make more concrete recommendations based on the results for future policymaking. At the same time, performing multiple case studies can make better recommendations on the effect of climate change on the structures or identify risks.

In terms of actions, it might be interesting to include multiple inspection actions or a 'bonus' for grouping maintenance actions with, for instance, a mobilization cost. Also, more detailed modeling of the effects of maintenance actions is recommended, which closer reassembles a real-life maintenance action. The limit state could be extended by including multiple failure mechanisms or modeling the dependency between components for failure. It is especially important that the effect of climate change plays a major part in the structural integrity of the structure which should follow from the risk analysis performed beforehand. Such a dependency can be in terms of the loading, e.g. extreme weather events or precipitation, or in terms of resistance, e.g. corrosion, or scour. An example of such a case study could be the monitoring and maintenance of dykes and banks in the Netherlands, which depend greatly on the state of the climate due to sea-level rise. The physical modeling in terms of climate change and deterioration has been quite limited within this study to narrow the scope of the research. This is evident in using a bilinear model to simulate the climate scenarios and using one deterioration mechanism in the SPD model. The added complexity of a comprehensive case study not only increases the effectiveness of a DRL algorithm compared to traditional heuristic-based approaches but is also closer to the reality of complex engineering systems.

6.6.2. Deep Reinforcement Learning

Deep reinforcement learning is another path that was not explored extensively in this research due to its limited scope. For the DRL part, recommendations are made based on the algorithm applied and the modeling of such an algorithm Regarding the DRL part of the work, it is recommended to look at different algorithms. Especially the use of DDMAC [45] is recommended when introducing a more complex case study with more components or actions. In DDMAC, the actor-network is decentralized, so each component has its neural network. This means each component has its agent, independently choosing actions for each component. Each agent still observes a centralized state of the environment and acts accordingly. It might also be interesting to apply DDMAC categorically, mapping similar components to one neural network to decrease the total network parameters and decrease the required training time. This is also in line with the results where it was seen that the policy tries to converge to a similar policy for like-wise components. Also, this seems like a logical way forward for a complex engineering system with multiple types of components. Different elements could be grouped into one network if an entire bridge were modeled. An example could be to divide a girder bridge into the pillars, the slab, and the girders. The results of such a model would be more intuitive, leading to similar policies for

like-wise components, adding a feeling of safety for the planner. A major concern is regarding the lack of transparency of many AI algorithms. The lack of transparency is due to the internal mechanism being hidden from view and decision-making being difficult to apprehend. This aligns with the statement in the implications, where the framework should be used as a tool for the engineer. Having a more intuitive I&M policy will make the interaction between the engineer and the algorithm easier.

The modeling of the algorithm contributes significantly to the final results gathered. The modeling consists of hyper-parameter tweaking during the training process, a sensitivity analysis of the model's key parameters, and the reward function's setup. Hyper-parameter tweaking can significantly impact the training time required to converge toward an optimal policy. A significant limitation of policies derived from DRL algorithms is its requirement for high computational resources. The thesis did not contain a sensitivity analysis of the critical parameters or different configurations of the reward function. Incorporating both when using the framework for an extensive case study is recommended. The sensitivity analysis is required to validate the results gathered and provide more reliable results. This is especially needed when making statements on the effect of climate change on our infrastructure or when one wants to adopt a policy derived from DRL to an existing infrastructure. The reward function is a critical component of deep reinforcement learning, providing a goal for the agent and providing feedback. The design of such a reward function is challenging and can significantly contribute to the final policy derived by the agent. The reward function also introduces bias into the agent's decision-making, leading to convergence towards a lousy policy. An example for the case study could be neglecting a risk-cost in the reward function, which would lead to the optimal policy being never to repair or inspect (since this results in the lowest lifecycle costs). Therefore, it is recommended to incorporate a significant portion of one's time into sensitivity analysis and multiple instances of a setup of the reward function if the framework is to be applied to an extended case study.

6.7. Conclusion

The discussion will be finalized in this section by forming a conclusion based on answering the research questions. This will be done by reiterating the key findings of the study. First, the sub-questions will be formulated and answered and used to form an answer to the main research question.

- **How does a traditional I&M plan compare to a dynamic policy proposed by the framework in a changing climate?**

  The traditional I&M plannings considered were TBM and CBM. TBM did not change its planning since the decision-making did not consider the damage state of the components due to the preventive maintenance approach. In CBM, the deteriorating structure's actual condition is assessed using thresholds on the results of inspections. Therefore, CBM is better suited as an I&M planning than TBM when considering a changing climate. CBM, however, does not make a distinction between the different components and dependencies present in an actual engineering structure. The dynamic policy proposed by the framework can take the partial observability of climate change into account in the decision-making and adapt its policy accordingly. This can be seen in figure 5.8, with PPO having different policies for the scenarios. Another important distinction is the ability of the framework to work with model uncertainty related to climate change, which the benchmarks cannot. The framework can adapt its policy dynamically based on the state of the climate, while benchmarks are optimized based on one scenario and, therefore, cannot.

- **What changes in the I&M actions are proposed by the dynamic policy of the framework under a changing climate, and are they significant?**

  The results from the framework show that different climate scenarios lead to other policies in decision-making for I&M planning. By using a belief over the current state of the climate, the agent can change its policy on the fly and take action based on the changing climate. Changes proposed by the framework are, for instance, an increase in the maintenance actions performed for scenarios with worse climate change and a distinction in the adopted policy for different components. It must, however, be noted that these results are not robust and do not mean that climate change equals more maintenance. To draw such conclusions, a more extensive case study, including sensitivity analysis and multiple instances of a reward function, must be performed, as

stated in the recommendations section. The results show that the dynamic policy can capture environmental changes and plan accordingly.

With a conclusion to the sub-questions found, let us return to the main research question:

What is the effect of a framework that incorporates model uncertainty in the deteriorating environment of reinforced concrete due to climate change in the inspection and maintenance plan of bridges to lower the lifecycle costs and maintain safety?

To answer this research question, the following approach was adopted. A theoretical understanding has been performed and presented in the literature review on the three main topics: climate change, deterioration of reinforced concrete, and optimal inspection and maintenance planning. The theoretical background was translated to a setup of the framework to incorporate partial observability of climate change within the decision-making of inspection and maintenance planning. Climate change is incorporated into the planning by having a belief state over the possible climate scenarios. The proposed framework is established in three phases: setup of the physical model, setup of the POMDP, and finding the optimal policy.

The framework is applied to a case study to verify the approach and draw conclusions on the effectiveness of the framework using quantitative data analysis. In this research, a system has been found that allows climate change to be incorporated into decision-making by using a belief over different climate scenarios. The study has found that the framework can adapt to different climate scenarios and plan accordingly, which can be seen by the other I&M policies provided in figure 5.8. The policy derived from the framework not only leads to lower life-cycle costs under each climate scenario but also maintains safety which can be seen by the comparison between risks provided in figure 5.16 and figure 5.17.

The study has also found that it is required to perform a more extensive case study to pinpoint the exact effects of such a framework compared to traditional approaches. This is especially the case due to the criticality of a well-established reward function for the reliability of such a policy. The reward function is dependent on the case study as well. An example was discussed over the dependency between the climate variables of the number of frost days and average annual temperature, presented in section 6.2.

Moreover, this research indicates that the use of Deep Reinforcement Learning to integrate the model uncertainties of climate scenarios into the decision-making of inspection and maintenance planning is the way to go. Especially with the ever-increasing computational power available to humankind, increases in the maintenance costs of our infrastructure, and the increases in the effects of climate change, such advanced policies will become a necessity. However, it is a new field of inspection and maintenance planning; therefore, similar studies should be performed to create a clear scientific foundation before applying it in real life.

References

[1] A.N. Bleijenberg. *Instandhouding civiele infrastructuur, Proeve van landelijk prognoserapport vervanging en renovatie*. Tech. rep. Delft: TNO, Apr. 2021.

[2] Günter a. Schmitt et al. “Global Needs for Knowledge Dissemination, Research, and Development in Materials Deterioration and Corrosion Control”. In: *World Corrosion Organization* May (2009).

[3] A Nasr et al. “A review of the potential impacts of climate change on the safety and performance of bridges”. In: *Sustainable and Resilient Infrastructure* 6.3-4 (2019), pp. 192–212. DOI: 10.1080/23789689.2019.1593003. URL: https://www.tandfonline.com/action/journalInformation?journalCode=tsri20.

[4] E Bastidas-Arteaga et al. “Influence of global warming on durability of corroding RC structures: A probabilistic approach”. In: *Engineering Structures* 51 (2013). ISSN: 01410296. DOI: 10.1016/j.engstruct.2013.01.006.

[5] X Wang et al. *Analysis of Climate Change Impacts on the Deterioration of Concrete Infrastructure - Part 1: Mechanisms, Practices, Modelling and Simulations-A review*. Tech. rep. 2010.

[6] Xiaoming Wang et al. “Analysis of Climate Change Impacts on the Deterioration of Concrete Infrastructure-Part 2: Modelling and Simulation of Deterioration and Adaptation Options”. In: (2010). URL: https://www.researchgate.net/publication/298646715.

[7] X Wang et al. *Analysis of Climate Change Impacts on the Deterioration of Concrete Infrastructures - Part 3 : Case Studies of Concrete Deterioration and Adaptation*. Tech. rep. 2010.

[8] C. P. Andriotis and K. G. Papakonstantinou. “Managing engineering systems with large state and action spaces through deep reinforcement learning”. In: *Reliability Engineering and System Safety* 191 (Nov. 2019). ISSN: 09518320. DOI: 10.1016/j.ress.2019.04.036.

[9] J. B. R. Matthews. “Annex I: Glossary”. In: *IPCC* (2018). DOI: 10.1017/9781009157940.008. URL: https://doi.org/10.1017/9781009157940.008..

[10] Valérie Masson-Delmotte et al. *Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Tech. rep. IPCC, 2021. URL: www.ipcc.ch.

[11] Detlef P. van Vuuren et al. “The representative concentration pathways: An overview”. In: *Climatic Change* 109.1 (Nov. 2011), pp. 5–31. ISSN: 01650009. DOI: 10.1007/S10584-011-0148-Z/TABLES/4. URL: https://link.springer.com/article/10.1007/s10584-011-0148-z.

[12] Intergovernmental Panel on Climate Change (IPCC). *Climate change 2014 : synthesis report. Constribution of Working Groups I, II, and III to the Fifth Assessment Report of the Intergovernmental Panel of Climate Change*. Tech. rep. Geneva: IPCC, 2014.

[13] KNMI. *KNMI ’14 climate scenario’s for the Netherlands: a guide for professionals in climate adaptation*. Tech. rep. De Bilt: KNMI, 2014.

[14] KNMI. *KNMI’14: Climate Change scenarios for the 21st Century-A Netherlands perspective - scientific Report WR2014-01*. Tech. rep. De Bilt: KNMI, 2014. URL: www.climatescenarios.nl.

[15] Lizhengli Peng and Mark G. Stewart. “Climate change and corrosion damage risks for reinforced concrete infrastructure in China”. In: *Structure and Infrastructure Engineering* 12.4 (2016). ISSN: 17448980. DOI: 10.1080/15732479.2013.858270.

[16] Emilio Bastidas-Arteaga and Mark G. Stewart. “Damage risks and economic assessment of climate adaptation strategies for design of new concrete structures subject to chloride-induced corrosion”. In: *Structural Safety* 52.PA (2015). ISSN: 01674730. DOI: 10.1016/j.strusafe.2014.10.005.

[17] Amro Nasr, Dániel Honfi, and Oskar Larsson Ivanov. “Probabilistic analysis of climate change impact on chloride-induced deterioration of reinforced concrete considering Nordic climate”. In: *J Infrastruct Preserv Resil* 3.8 (2022). DOI: 10.1186/s43065-022-00053-6. URL: https://doi.org/10.1186/s43065-022-00053-6.

[18] S Talukdar. *The effects of global climate change on carbonation induced corrosion of reinforced concrete structures*. Tech. rep. Vancouver: University of British Columbia, Apr. 2013.

[19] Dimitri V. Val, Mark G. Stewart, and Robert E. Melchers. “Effect of reinforcement corrosion on reliability of highway bridges”. In: *Engineering Structures* 20.11 (Nov. 1998), pp. 1010–1019. ISSN: 0141-0296. DOI: 10.1016/S0141-0296(97)00197-1.

[20] A.M. Neville. *Properties of Concrete: Properties of Concrete, 5th Edition*. Ed. by Pearson Publishing. 5th edition. 2011. ISBN: 978-0-273-75580-7.

[21] Guofang Chen et al. “Carbonation depth predictions in concrete structures under changing climate condition in China”. In: *Engineering Failure Analysis* 119 (2021). ISSN: 13506307. DOI: 10.1016/j.engfailanal.2020.104990.

[22] Yihui Zhou et al. “Carbonation-Induced and Chloride-Induced Corrosion in Reinforced Concrete Structures”. In: (2014). DOI: 10.1061/(ASCE)MT.1943-5533.0001209. URL: https://www.researchgate.net/publication/268924426.

[23] Mark G Stewart and David V Rosowsky. “Time-dependent reliability of deteriorating reinforced concrete bridge decks”. In: *Structural Safety* 20 (1998), pp. 91–109.

[24] DuraCrete. *Probabilistic Performance based Durability Design of COncrete Structures*. Tech. rep. The European Union - Brite EuRam III, 2000.

[25] Ueli M. Angst. “Predicting the time to corrosion initiation in reinforced concrete structures exposed to chlorides”. In: *Cement and Concrete Research* 115 (Jan. 2019), pp. 559–567. ISSN: 00088846. DOI: 10.1016/J.CEMCONRES.2018.08.007.

[26] Dimitri V. Val and Pavel A. Trapper. “Probabilistic evaluation of initiation time of chloride-induced corrosion”. In: *Reliability Engineering & System Safety* 93.3 (Mar. 2008), pp. 364–372. ISSN: 0951-8320. DOI: 10.1016/J.RESS.2006.12.010.

[27] Rijkswaterstaat. *Inventarisatie Kunstwerken (bruggen, tunnels en viaducten)*. Tech. rep. Ministerie van Verkeer en Waterstaat, Sept. 2007.

[28] Zhao Hui Lu et al. “Empirical model of corrosion rate for steel reinforced concrete structures in chloride-laden environments”. In: *Advances in Structural Engineering* 22.1 (Jan. 2019), pp. 223–239. ISSN: 20484011. DOI: 10.1177/1369433218783313.

[29] CBS et al. *Meteorologische gegevens, 1990-2017 | Environmental Data Compendium*. Nov. 2018. URL: https://www.clo.nl/en/node/26967.

[30] Dimitri V Val. “Deterioration of Strength of RC Beams due to Corrosion and Its Influence on Beam Reliability”. In: *Journal of Structural Engineering* 133 (2007), pp. 1297–1306. DOI: 10.1061/ASCE0733-94452007133:91297.

[31] Dimitri V Val and Robert E Melchers. “Reliability of deteriorating RC slab bridges”. In: *Journal of Structural Engineering* 123 (1997), pp. 1638–1644.

[32] Amin Jamali et al. “Modeling of corrosion-induced concrete cover cracking: A critical analysis”. In: *Construction and Building Materials* 42 (May 2013), pp. 225–237. ISSN: 0950-0618. DOI: 10.1016/J.CONBUILDMAT.2013.01.019.

[33] J. Rodriguez, L. M Ortega, and J. Casal. “Corrosion of reinforcement and service life of concrete structures”. In: *Proceeding of the 7th international conference on durability of building materials and components* (1996), pp. 117–126.

[34] J.A. Mullard and M.G. Stewart. “Corrosion-induced cracking: New test data and predictive models”. In: *ACI Structural journal* 108.1 (2011).

[35] J M Van Noortwijk and M D Pandey. “A stochastic deterioration process for time-dependent reliability analysis”. In: (2004).

[36] J. M. van Noortwijk. “A survey of the application of gamma processes in maintenance”. In: *Reliability Engineering & System Safety* 94.1 (Jan. 2009), pp. 2–21. ISSN: 0951-8320. DOI: 10.1016/J.RESS.2007.03.019.

[37] Shiyin Wei, Yuequan Bao, and Hui Li. “Optimal policy for structure maintenance: A deep reinforcement learning framework”. In: *Structural Safety* 83 (Mar. 2020), p. 101906. ISSN: 0167-4730. DOI: 10.1016/J.STRUSAFE.2019.101906.

[38] Rijkswaterstaat. *Factsheet inspectie van een brug*. 2015. URL: https://puc.overheid.nl/rijkswaterstaat/doc/PUC_162237_31/.

[39] Pablo G. Morato et al. “Inference and dynamic decision-making for deteriorating systems with probabilistic dependencies through Bayesian networks and deep reinforcement learning”. In: *Reliability Engineering and System Safety* (Sept. 2022). DOI: 10.48550/arxiv.2209.01092. URL: https://arxiv.org/abs/2209.01092v1.

[40] K. G. Papakonstantinou and M. Shinozuka. “Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part I: Theory”. In: *Reliability Engineering & System Safety* 130 (Oct. 2014), pp. 202–213. ISSN: 0951-8320. DOI: 10.1016/J.RESS.2014.04.005.

[41] Richard S Sutton and Andrew G Barto. *Reinforcement Learning: An Introduction Second edition, in progress*. Second. London: The MIT Press, 2014.

[42] K. G. Papakonstantinou, C. P. Andriotis, and M. Shinozuka. “POMDP and MOMDP solutions for structural life-cycle cost minimization under partial and mixed observability”. In: *Structure and Infrastructure Engineering* 14.7 (July 2018), pp. 869–882. ISSN: 17448980. DOI: 10.1080/15732479.2018.1439973. URL: https://www.tandfonline.com/doi/abs/10.1080/15732479.2018.1439973.

[43] K. G. Papakonstantinou and M. Shinozuka. “Optimum inspection and maintenance policies for corroded structures using partially observable Markov decision processes and stochastic, physically based models”. In: *Probabilistic Engineering Mechanics* 37 (2014), pp. 93–108. ISSN: 02668920. DOI: 10.1016/J.PROBENGMECH.2014.06.002.

[44] K. G. Papakonstantinou, C. Andriotis, and M. Shinozuka. “Point-based pomdp solvers for life-cycle cost minimization of deteriorating structures”. In: *Life-Cycle of Engineering Systems: Emphasis on Sustainable Civil Infrastructure - 5th International Symposium on Life-Cycle Engineering, IALCCE 2016* (Nov. 2017), pp. 427–434. DOI: 10.1201/9781315375175-56. URL: https://www.taylorfrancis.com/chapters/edit/10.1201/9781315375175-56/point-based-pomdp-solvers-life-cycle-cost-minimization-deteriorating-structures-papakonstantinou-andriotis-shinozuka.

[45] C. P. Andriotis and K. G. Papakonstantinou. “Deep reinforcement learning driven inspection and maintenance planning under incomplete information and constraints”. In: *Reliability Engineering and System Safety* 212 (July 2020). ISSN: 09518320. DOI: 10.48550/arxiv.2007.01380. URL: https://arxiv.org/abs/2007.01380v1.

[46] Y Li. “Deep Reinforcement Learning”. In: *Smart Innovation, Systems and Technologies* 273 (Oct. 2018), pp. 136–142. ISSN: 21903026. DOI: 10.48550/arxiv.1810.06339. URL: https://arxiv.org/abs/1810.06339v1.

[47] Kai Arulkumaran et al. “Deep reinforcement learning: A brief survey”. In: *IEEE Signal Processing Magazine* 34.6 (Nov. 2017), pp. 26–38. ISSN: 10535888. DOI: 10.1109/MSP.2017.2743240.

[48] Volodymyr Mnih et al. *Playing Atari with Deep Reinforcement Learning*. Tech. rep. 2013.

[49] Hado Van Hasselt, Arthur Guez, and David Silver. “Deep Reinforcement Learning with Double Q-Learning”. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 30.1 (Mar. 2016), pp. 2094–2100. ISSN: 2374-3468. DOI: 10.1609/AAAI.V30I1.10295. URL: https://ojs.aaai.org/index.php/AAAI/article/view/10295.

[50] Richard S. Sutton et al. “Policy Gradient Methods for Reinforcement Learning with Function Approximation”. In: *Advances in Neural Information Processing Systems* 12 (1999).

[51] Ronald J. Williams. “Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning”. In: *Machine Learning* 8 (1992), pp. 229–256. DOI: 10.1007/978-1-4615-3618-5{_}2. URL: https://link.springer.com/chapter/10.1007/978-1-4615-3618-5_2.

[52] Volodymyr Mnih et al. “Asynchronous Methods for Deep Reinforcement Learning”. In: *33rd International Conference on Machine Learning, ICML 2016* 4 (Feb. 2016), pp. 2850–2869. DOI: 10.48550/arxiv.1602.01783. URL: https://arxiv.org/abs/1602.01783v2.

[53] John Schulman et al. *Trust Region Policy Optimization*. June 2015. URL: https://proceedings.mlr.press/v37/schulman15.html.

[54] John Schulman et al. *Proximal Policy Optimization Algorithms*. Tech. rep. 2017.

[55] C Lathourakis. *Optimal maintenance of deteriorating systems integrating deep reinforcement learning and Bayesian inference*. Tech. rep. TU Delft, 2022.

[56] Tom Rupke. *Framework for a Risk Informed Maintenance Strategy for Deteriorating Hydraulic Structures*. 2013.

[57] Nederlands Normalisatie-instituut. *Eurocode 1: Belastingen op constructies-Deel 2: Verkeersbelasting op bruggen Eurocode 1: Actions on structures-Part 2: Traffic loads on bridges*. Tech. rep. Delft, 2015.

[58] Nederlands Normalisatie-instituut. *NEN-EN 1990+A1+A1/C2, Eurocode: Grondslagen voor het constructief ontwerp*. Tech. rep. 2015.

[59] Sher Ali Mirza, Michael Hatzinikolas, and James G. MacGregor. “Statistical Descriptions of Strength of Concrete”. In: *ASCE J Struct Div* 105.6 (1979). ISSN: 0044-8001. DOI: 10.1061/jsdeag.0005161.

[60] Centraal Bureau voor de Statistiek (CBS). *Gemiddelde temperatuur*. 2017. URL: https://www.cbs.nl/nl-nl/maatschappij/natuur-en-milieu/groene-groei/milieukwaliteit-van-het-leven/indicatoren/gemiddelde-temperatuur.

[61] Greg Brockman et al. *OpenAI Gym*. 2016.

[62] Antonin Raffin et al. “Stable-Baselines3: Reliable Reinforcement Learning Implementations”. In: *Journal of Machine Learning Research* 22.268 (2021), pp. 1–8.

[63] OpenAI. *Proximal Policy Optimization — Spinning Up documentation*. 2018. URL: https://spinningup.openai.com/en/latest/algorithms/ppo.html.

[64] Elizabeth Bismut and Daniel Straub. “Optimal adaptive inspection and maintenance planning for deteriorating structural systems”. In: *Reliability Engineering & System Safety* 215 (Nov. 2021), p. 107891. ISSN: 0951-8320. DOI: 10.1016/J.RESS.2021.107891.

[65] Delft High Performance Computing Centre (DHCP). *Delft Blue Supercomputer (Phase 1)*. 2022.

[66] Andrzej S. Nowak and Young Kyun Hong. “Bridge Live Load Models”. In: *Journal of Structural Engineering* 117.9 (1991). ISSN: 0733-9445. DOI: 10.1061/(asce)0733-9445(1991)117:9(2757).

[67] S N Jonkman et al. *Probabilistic Design: Risk and Reliability Analysis in Civil Engineering Lecture notes CIE4130*. 2018.

A

The design of the bridge

This appendix covers the initial design of the case study covered in chapter 4 according to the loading of Eurocode 1 part 2 and the reliability of Eurocode 0. First, the loading of the case study will be introduced and a translation to limit states for each component is made. Next, the initial design of the reinforcement of the structure is calculated using a Monte Carlo analysis to show that the initial reliability corresponds to Eurocode. The appendix is finalized with a check to show that no shear reinforcement is required.

A.1. Loading

The bridge is loaded by a combination of dead loads and a live truckload. The self-weight of the reinforced concrete deck has a stochastic distribution of mean to nominal (G/G_n) of 1.05 with a COV of 0.10 [23]. The live load consists of the weight of a truck with a dynamic amplification factor plus a distributed load for general traffic in compliance with Eurocode 1 part 2: Traffic loads on bridges [57].

The distribution of the weight of the truck is described according to a survey of trucks in the USA which found a mean truck load of 250 kN and a COV of 0.4 [66]. The mean dynamic load factor is 1.15, which follows from the Eurocode 1 part 2 [57]. The truck is assumed to be a combination of two concentrated loads working on the bridge. The axle spacing and distribution of the force are taken from the two-axis truck given in table 4.7 from the Eurocode [57] and summarized in figure 4.1. It has been found that the critical annual load can be simulated when two fully correlated trucks are side by side. For n-number of fully correlated trucks per year, with a normally distributed truck weight, the annual heaviest truck weight can be sampled from the following cumulative distribution function [23]:

$$F_n(w) = \left[\Phi\left(\frac{w-\mu_w}{\sigma_w}\right)\right]^n \tag{A.1}$$

with: Φ is the cumulative distribution function of the standard normal distribution, μ_w the mean weight of the truck, and σ_w the standard deviation of the truck weight.

It is assumed that 600 trucks pass over the bridge annually in compliance with [23]. The distributed load for general traffic follows from the Eurocode [57] with a deterministic value of $9kN/m^2$ acting on a width of 3 meters for the heaviest loaded traffic lane. The loading has been summarized in table A.1 below.

Parameter	Mean	COV	Unit	Distribution	Reference
G	$G_{nom} \cdot 1.05$	0.10	$[kN/m^3]$	Lognormal	[23]
Q	9		$[kN/m^2]$	Deterministic	[57]
F	287.5	0.41	$[kN]$	Normal	[23]

Table A.1: Loading on the structure

The loading is modeled stochastically where different components have different linear combinations of the loading to maximize the flexural stresses at the critical cross-sections.

From the linear static analysis, the following linear combinations for the different cross-sections are considered:

- $M_{S,1}, M_{S,5} = 1.70F + 6.68(G + Q)$
- $M_{S,2}, M_{S,4} = \min\left[-1.57F - 11.88(G + Q), -1.53F - 12.40(G + Q)\right]$
- $M_{S,3} = 1.53F + 6.20(G + Q)$

A.2. Initial design

For the initial design, the input parameters of tables 4.1 - 4.6 are used. The applied loading is given in the section above. The probability of failure of the initial design is calculated using a Monte Carlo analysis with importance sampling. The initial annual failure probability according to Eurocode 0 [58] should be $P_{F,0} = 1.3 \cdot 10^{-6}$. This low failure probability requires many samples to be drawn from the initial limit state to approximate the failure probability which requires a high computational time. Therefore a method called importance sampling is applied which transforms a stochastic distribution with a sampling distribution to obtain more samples in the unsafe domain and transforms it back towards the original distribution afterward. More information about importance sampling can be found in [67].

The limit state is given per component and is described below:

$$Z_i = M_{R,i} - M_{S,i} \tag{A.2}$$

The resisting bending moment $M_{R,i}$ consists of the internal lever arm z, the yield strength of the steel f_y and the cross-sectional area of the reinforcement $A_{s,i}$.

$$M_{R,i} = A_{s,i} \cdot f_y \cdot z \tag{A.3}$$

where the lever arm z can be determined according to equation A.4, and equation A.5 for concrete classes $\leq$ C50/60.

$$x_u = \frac{A_{s,i} * f_y}{0.75 * f_c * width} \tag{A.4}$$

$$z = d - 0.39 * x_u \tag{A.5}$$

Where the failure probability per component $p_i = P(Z_i < 0)$ is used to compute the failure probability of the entire structure under the assumption of independence between the components.

$$P_F = 1 - \prod_{i=0}^{5}(1 - p_i) \tag{A.6}$$

The mean of the yield strength of the reinforcement f_y has been shifted toward the failure domain. The original distribution of the yield strength is $f_y \sim LN(\mu = 490, \sigma = 49)$ and the sampling distribution is $f_y s \sim LN(\mu = 290, \sigma = 49)$. To show that the importance sampling approximates the exact failure probability it is checked for a design with a higher initial failure probability with crude Monte Carlo. This has been done by applying a lower initial reinforcement ratio as shown below.

- Extra bottom reinforcement for first span: $A_{s1} = \varnothing 16 - 150$
- Extra top reinforcement at supports: $A_{s2} = \varnothing 16 - 100$
- Extra bottom reinforcement for middle span: $A_{s3} = \varnothing 16 - 150$

Then 100,000 samples are drawn for both the crude Monte-Carlo analysis and Monte-Carlo analysis with importance sampling. This is done for 100 iterations and the average of the iterations is given below. The results show that the importance sampling finds the right failure probability and will therefore be used from here on.

- Probability of failure according to Crued Monte Carlo: $P_{F,0} = 0.0645$
- Probability of failure according to importance sampling: $P_{F,0} = 0.0645$

Next, the initial design of the reinforcement is calculated with the following assumptions:

- The probability of failure of the initial design at $t = 0$ should be lower than $P_{F,0} = 1.3 \cdot 10^{-6}$.
- The probability of failure of each component p_i should be approximately equal.
- The same reinforcement diameter $\varnothing = 20mm$ is used but with different spacing of the reinforcement.
- For 10 iterations, 1,000,000 samples are drawn, and the average value is used.

This leads to the following design:

- $A_{s1} = \varnothing 20 - 140mm$ with a failure probability of the component $p_1 = 1.633 \cdot 10{-}7$
- $A_{s2} = \varnothing 20 - 105mm$ with a failure probability of the component $p_2 = 2.199 \cdot 10^{-7}$
- $A_{s3} = \varnothing 20 - 160mm$ with a failure probability of the component $p_3 = 1.765 \cdot 10^{-7}$
- Initial annual failure probability: $P_{F,0} = 0.90 * 10^{-6}$

A.3. Shearforce

The bridge does not require shear reinforcement according to the minimum shear resistance check from Eurocode 2 (6.2.b) [57]. The width of the bridge is 11m. There are two driving lanes, with two driving directions.
It is assumed that shear is not governing the design. To make this assumption work, a calculation has been performed according to Eurocode 2 6.2.2 to prove that the slab complies with the minimum rule that no shear reinforcement is required.

$$V_{rd,c} = [C_{Rd,c}k(100\rho_1 f_{ck})^{1/3} + k_1\sigma_{cp}]b_w d \tag{A.7}$$

with a minimum of:

$$V_{rd,c} = (v_{min} + k_1 * \sigma_{cp}) * b_w * d \tag{A.8}$$

$$v_{min} = 0.035k^{3/2}\sqrt{f_{ck}} \tag{A.9}$$

$$k = 1 + \sqrt{200/d} \leq 2.0 \tag{A.10}$$

If half of the bridge is considered ($b_w = 5500mm$) and no normal stresses are present in the cross-section $\sigma_{cp} = 0$ and a characteristic concrete strength of C30/37 = $f_{ck} = 30N/mm^2$ and a height of $h = 532mm -> d = 435mm$.

$$k = 1 + \sqrt{200/435} = 1.68 \leq 2.0 \tag{A.11}$$

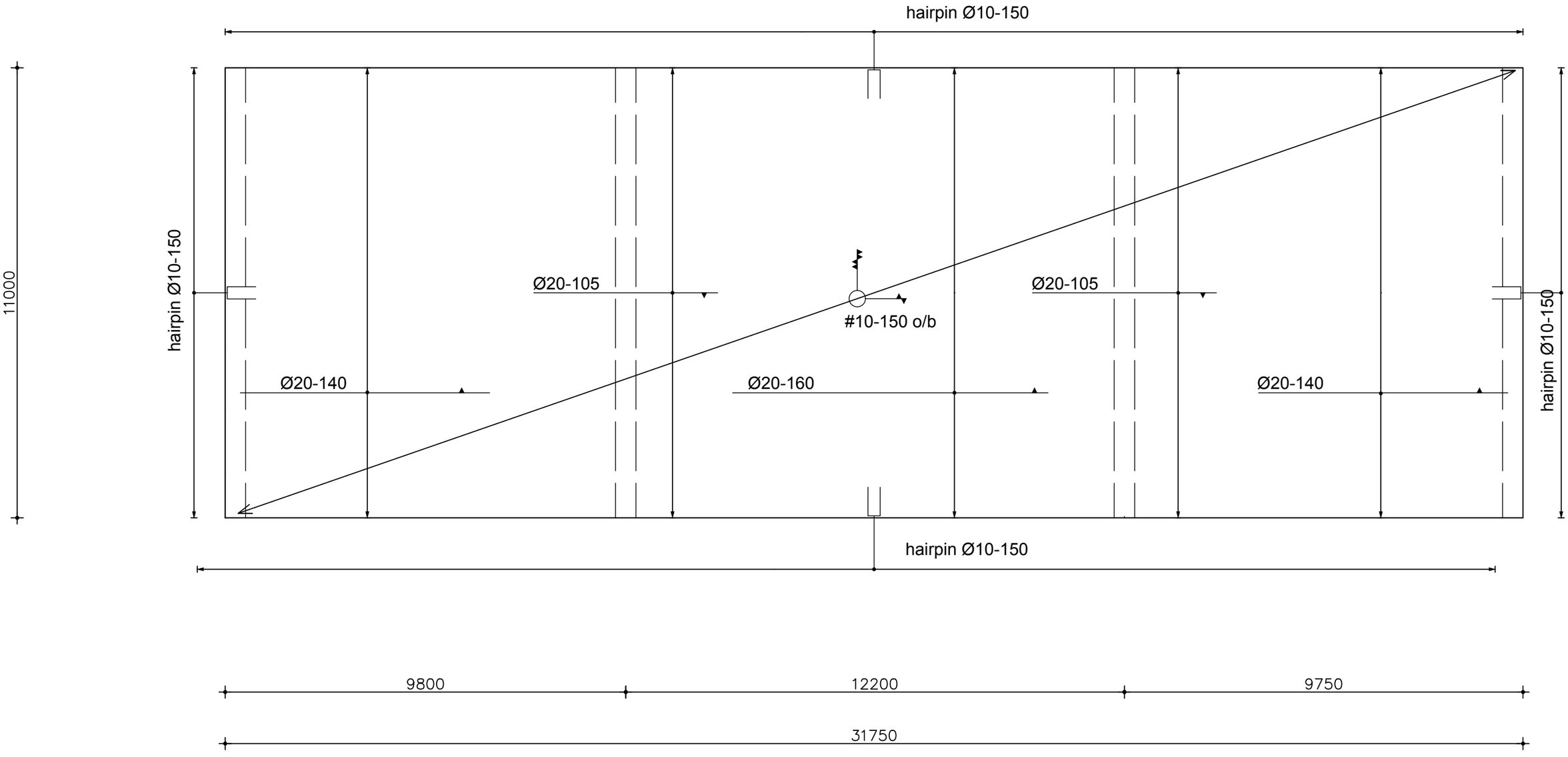
$$v_{min} = 0.035 * 1.68^{3/2}\sqrt{30} = 0.42 \tag{A.12}$$

$$V_{rd,c} = 0.42 \cdot 5500 \cdot 435 = 997kN \tag{A.13}$$

According to the linear static analysis performed in Matrix, the design value of the shear force $V_{ed} < V_{rd,c}$, so no shear reinforcement is required.

B

Illustrations of the case study



Reinforcement drawing of the RC slab.

Scale: 1/100

Annotation reinforcement layer
First layer top reinforcement
Second layer top reinforcement
Second layer bottom reinforcement
First layer bottom reinforcement