Titre: Title:	Damage Detection for Structural Health Monitoring Using Reinforcement and Imitation Learning
Auteur: Author:	Shervin Khazaeli
Date:	2022
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Khazaeli, S. (2022). Damage Detection for Structural Health Monitoring Using Reinforcement and Imitation Learning [Thèse de doctorat, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/10744/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/10744/
Directeurs de recherche: Advisors:	James Alexandre Goulet
Programme: Program:	Génie civil

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Damage Detection for Structural Health Monitoring
Using Reinforcement and Imitation Learning**

SHERVIN KHAZAELI

Département de génie civil, géologique et des mines

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
génie civil

Novembre 2022

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Cette thèse intitulée :

**Damage Detection for Structural Health Monitoring
Using Reinforcement and Imitation Learning**

présentée par **Shervin KHAZAELI**
en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

Nicolas SAUNIER, président
James GOULET, membre et directeur de recherche
Raimondo BETTI, membre
Benjamin MIQUEL, membre externe

DEDICATION

To my loving family & individuals who have always inspired me.

ACKNOWLEDGEMENTS

RÉSUMÉ

La détérioration et la dégradation des infrastructures civiles telles que les ponts sont des problèmes communs à la plupart des pays industrialisés. Ils entraînent des *dommages structurels* qui sont à l'origine de dépenses importantes liées au maintien de la sécurité et de l'état de fonctionnement des infrastructures. Dans le contexte de la surveillance de la santé structurelle (SHM), une approche courante consiste à utiliser des *capteurs* afin de surveiller l'état des structures, par exemple, normal ou anormal, et d'identifier la présence de dommages structurels. La détection des dommages à l'aide d'un système SHM est une tâche difficile car ils se développent souvent au fil des années et ils ont généralement un effet sur les réponses structurelles qui est d'un ordre de grandeur inférieur aux effets externes tels que la température. En plus de détecter les anomalies, nous nous intéressons à quantifier la capacité à détecter à juste titre ces dommages, c'est-à-dire *quantifier la détectabilité des dommages structurels*, afin d'évaluer la viabilité et l'utilité d'un système SHM. Lorsqu'un dommage structurel se produit, les réponses structurelles passent d'un état normal à un état anormal. Ici, une *anomalie* fait référence à un tel écart, et est quantifiée par rapport à ses caractéristiques consistant en l'ampleur de l'anomalie et sa durée. Ces caractéristiques représentent la gravité d'un dommage structurel et son temps de développement, respectivement. Par conséquent, les principaux objectifs d'un système SHM sont de détecter les changements dans les réponses structurelles, c'est-à-dire les anomalies, et en outre, d'associer l'ampleur de ces changements à la gravité des dommages structurels.

Les méthodologies de détection d'anomalies existantes ne sont généralement pas en mesure d'associer l'apparition d'anomalies à des dommages structurels; ces méthodes manquent d'un mécanisme pour quantifier la probabilité de détecter des anomalies en fonction de l'ampleur d'une anomalie, de sa durée et du temps nécessaire pour détecter une anomalie après son début, c'est-à-dire *quantifier la détectabilité de l'anomalie*. De plus, la capacité des méthodes d'interprétation des données existantes utilisées pour détecter les anomalies est limitée : ces procédés ne sont pas en mesure de faire la distinction de manière fiable entre un changement dû à un écart par rapport à la condition normale, c'est-à-dire une anomalie, et un changement dû à des variations d'effets externes, par exemple les fluctuations de température saisonnières. En conséquence, les méthodes existantes génèrent souvent un grand nombre d'alarmes fausses qui compromettent la viabilité d'un système de surveillance.

Cette thèse présente une nouvelle approche de détection d'anomalies qui permet de quantifier de manière probabiliste la détectabilité des dommages structuraux. À cette fin, de

nouvelles méthodologies basées sur les données sont proposées afin de a) détecter les anomalies, b) quantifier de manière probabiliste la détectabilité des anomalies, c) associer l'ampleur des anomalies aux niveaux de gravité des dommages structurels, et d) mesurer la capacité d'une configuration de capteurs pour distinguer les dommages structurels, c'est-à-dire la *distinguabilité des dommages.* Les méthodes de détection d'anomalies proposées dans ce travail de recherche reposent sur des modèles linéaires dynamiques bayésiens ainsi que sur l'apprentissage par renforcement et par imitation. Le premier permet de séparer les changements dans les réponses structurelles de ceux causés par des effets externes. Les deux derniers permettent d'incorporer les informations obtenues à partir des changements dans les réponses structurelles afin de détecter un large éventail d'anomalies avec différentes amplitudes et durées. De plus, la détectabilité des anomalies de chaque capteur est quantifiée à l'aide du taux annuel de fausses détections ainsi que de la probabilité de leurs vraies détections. Le premier est le nombre d'anomalies par an détectées à tort. Cette dernière est la probabilité de détecter correctement une anomalie par rapport à l'ampleur de l'anomalie, sa durée et le temps de détection. De plus, cette thèse propose une méthode afin d'examiner la capacité d'un capteur à détecter des dommages structurels avec différents niveaux de gravité. Cette méthode repose sur a) la quantification des changements dans les réponses structurelles causées par un dommage structurel basé sur un modèle physique de la structure, et b) la comparaison de ces changements à l'emplacement des capteurs avec les magnitudes de l'anomalie détectable obtenues à partir de la détectabilité de l'anomalie de chaque capteur. De plus, cette thèse développe une métrique afin de mesurer la distinguabilité des dommages de différentes configurations de capteurs, et par la suite de identifier la configuration qui maximise la distinguabilité des dommages. Les méthodologies proposées dans cette thèse sont validées à l'aide de mesures réelles recueillies sur trois travées instrumentées du pont Jacques-Cartier, Canada.

ABSTRACT

The deterioration and degradation of civil infrastructures such as bridges are common issues among most industrialized countries. They result in *structural damages* that are responsible for important expenses associated with maintaining the safety and serviceability of infrastructures. In the context of structural health monitoring (SHM), a common approach is to use *sensors* in order to monitor the condition of a structure, e.g., normal or abnormal, and identify the presence of structural damages. Detecting damages using a SHM system is a difficult task because they often develop over the span of years, and they typically have an effect on structural responses which is orders of magnitudes smaller than external effects such as the temperature. In addition to detecting anomalies, we are interested in quantifying the capacity to rightfully detect these damages, i.e., *quantifying the structural damage detectability*, in order to evaluate the viability and usefulness of a SHM system. When a structural damage occurs, structural responses depart from a normal condition to an abnormal one. Here, an *anomaly* is referred to such a departure, and is quantified with respect to its characteristics consisting in an anomaly's magnitude and its duration. These characteristics represent the severity of a structural damage and its time to develop, respectively. Therefore, the core objectives for a SHM system is to detect changes in structural responses, i.e., anomalies, and further, to associate the magnitudes of these changes to the severity of structural damages.

Existing anomaly detection methodologies are typically not able to associate the occurrence of anomalies to structural damages; These methods lack a mechanism to quantify the probability of detecting anomalies as a function of an anomaly's magnitude, its duration, and the time required to detect an anomaly after its starting time, i.e., *quantifying the anomaly detectability*. The capacity of existing data interpretation methods used for detecting anomalies are limited: These methods are not able to reliably discriminate between a change due to a departure from the normal condition, i.e., an anomaly, and a change due to external effect variations, e.g., the seasonal temperature fluctuations. As a result, existing methods often yield a large number of false alarms that undermine the viability of a monitoring system.

This thesis presents a new anomaly detection approach that enables quantifying probabilistically the structural damage detectability. To this end, new data-driven methodologies are proposed in order to a) detect anomalies, b) quantify probabilistically the anomaly detectability, c) associate the magnitudes of anomalies to the severity levels of structural damages, and d) measure the capacity of a sensor configuration in distinguishing structural damages, i.e., the *damage distinguishability*. The anomaly detection methods proposed in this thesis rely on

Bayesian dynamic linear models as well as the reinforcement and imitation learning methods. The former allows separating the changes in the structural responses from the ones caused by external effects. The latter two enable incorporating information obtained from the changes in the structural responses in order to detect a wide range of anomalies with different magnitudes and durations. Furthermore, the anomaly detectability of each sensor is quantified using the false detection rate of anomalies as well as the probability of their true detections. The former is the number of anomalies per year that are wrongfully detected. The latter is the probability of rightfully detecting an anomaly with respect to an anomaly's magnitude, its duration, and the time to required to detect an anomaly after its starting time. Moreover, this thesis proposes a method in order to examine the capacity of sensors to detect structural damages with different severity levels. This method relies on a) quantifying the changes in structural responses caused by a structural damage based on a physics-based model of the structure, and b) comparing these changes at the location of the sensors with the detectable anomaly's magnitudes obtained from the anomaly detectability of each sensor. Furthermore, this thesis develops a metric in order to measure the damage distinguishability of different sensor configurations, and subsequently identify the configuration that maximizes the damage distinguishability. The methodologies proposed in this thesis are validated using real measurements collected on three instrumented spans of the Jacques Cartier Bridge, Canada.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ACRONYMS

Symbols

$x^{\mathtt{AR}}$	Autoregressive hidden state variable
$\mathcal{A}$	Action set
$\mathcal{J}$	Anomaly function characteristics set
a	Agent's action
s	Agent's state
q_π	Action-value function
t_s	Anomaly function beginning time
t_e	Anomaly function end time
t_c	Anomaly function center time
f_a	Anomaly function
k	Anomaly function steepness
h_a	Anomaly function magnitude
w_a	Anomaly function duration
ϵ_a	Anomaly function tolerance
$x^{\mathtt{B}}$	Baseline hidden state variable
$\dot{x}^{\mathtt{B}}$	Baseline rate of change hidden state variable (speed)
$\boldsymbol{\Sigma} \equiv \mathrm{cov}$	Covariance matrix
$\mathbf{P}$	Cosine similarity matrix
$\mathcal{D}$	Data set
$\hat{\mathcal{Z}}$	Demonstration
Ξ	Demonstrations set
$\mathtt{M}, \mathtt{N}$	Discretization size
d	Discretization map
$\hat{\mathbf{d}}$	Damage vector
$\boldsymbol{\mu} \equiv \mathbb{E}$	Expected value vector
$\mathcal{Z}$	Episode
$\mathtt{E}$	Extensometer
c_ϵ	Exploration constant
$\mathbf{x}$	Hidden state variable vector
$\mathbf{r}_t$	Innovation vector
$\mathbf{G}_t$	Innovation covariance matrix

$\mathbf{M}$	Joint regime probability
$\mathbf{K}_t$	Kalman gain matrix
$x^{\mathtt{KR}}$	Kernel regression hidden state variable
ℓ	Kernel length or the initial length of the extensometer
$x^{\mathtt{L}}$	Level hidden state variable
$x^{\mathtt{LA}}$	Local acceleration hidden state variable
$\mathtt{I_L}$	Lateral inclinometer
$\mathtt{I_1}$	Longitudinal inclinometer on axis 1
$\mathtt{I_3}$	Longitudinal inclinometer on axis 3
c_α	Learning-rate constant
$\mathcal{M}$	Model matrices set
$\mathcal{L}_t$	Marginal likelihood
$\mathbf{d}$	Modified damage vector
$\mathtt{N}$	Number of visited states by an agent
$\mathtt{K}$	Number of consecutive episodes
$\mathtt{Q}$	Number of damage scenario
$\mathbf{y}$	Observation vector
$\mathbf{v}$	Observation model error vector
$\mathbf{C}$	Observation matrix
$\mathbf{R}$	Observation model error covariance matrix
p	Probability density or transition probability function
$\hat{\mathbf{y}}_t$	Predicted observation vector
$\pi^{\mathtt{s}}$	Probability of a stationary regime
$\pi^{\mathtt{ns}}$	Probability of a non-stationary regime
$\mathcal{P}$	Parameter set
π	Policy
Pr	Probability
Q	Q-value function
r	Reward function
β	Regression coefficient
$\ddot{x}^{\mathtt{B}}$	Speed rate of change hidden state variable (acceleration)
$\mathcal{S}$	State space (environment)
t	Time
$\mathbf{w}$	Transition model error vector
$\mathbf{A}$	Transition matrix
$\mathbf{Q}$	Transition model error covariance matrix

Δt	Time step between two consecutive times stamps
$\mathbf{Z}$	Transition matrix
$x^{\mathtt{T}}$	Trend hidden state variable
$\mathtt{T}$	Thermometer
$\mathcal{D}^{\mathrm{train}}$	Training data set
$\mathcal{D}^{\mathrm{test}}$	Test data set
$\mathbb{N}_{t>t_s}$	Time to detection
$\mathbb{U}$	Utility
ζ	Annual false alarm ratio
ϕ	Autoregressive coefficient
δ	Convergence error ratio
φ	Damage scenario set
γ	Discount factor
α	Learning rate
ϱ	Sensors set
σ	Standard deviation
ε	User-defined tolerance

Abbreviation

BDLM	Bayesian Dynamic Linear Model
BDL	Bearing Device Lockage
COR	Corrosion
FN	False Negative
FP	False Positive
KF	Kalman Filter
NU	Non-Uniform
PST	Partial Settlement
RL	Reinforcement Learning
SKF	Switching Kalman Filter
TN	True Negative
TP	True Positive
U	Uniform

LIST OF APPENDICES

CHAPTER 1   Introduction

The deterioration and degradation [7] of civil infrastructures such as bridges refer to the changes in the structural performance due to changes in the material or geometry of an structure. They are common issues among most industrialized countries where they are responsible for societal expenses affecting economical growth [8]. Civil infrastructures have to be maintained and monitored in order to ensure their serviceability and safety [9, 10]. To this end, *structural health monitoring* (SHM) systems have been of interest to many researchers and asset owners [11]. SHM systems often use *sensors* to provide continuous measurements about the structural responses and its surrounding environment [12]. These measurements are then interpreted in order to provide information about the state of a structure, e.g., normal or abnormal, as well as the presence of structural damages.

Structural responses depart from a normal condition to an abnormal one, when a structural damage occurs. Here, an *anomaly* refers to such a departure, and is quantified with respect to its characteristics consisting in an anomaly's magnitude and its duration. They represent the severity of a structural damage and its time to develop, respectively. The core objectives for a successful SHM system is to detect changes in structural responses, i.e., anomalies, and further to associate their magnitude to the severity levels of structural damages. Addressing these objectives involves coping with some challenges, because structural damages typically develop over the span of years, and they typically have an effect on structural responses which is orders of magnitudes smaller than external effects such as the temperature. Therefore, methodologies developed for interpreting data and detecting anomalies should be able to discriminate between changes in the structural responses from the external effects. In addition, SHM data interpretation and anomaly detection methods require to minimize the false detection of anomalies, maximize their true detections, and distinguish between different structural damages [11, 13]. Addressing these requirements would ensure a reliable and useful SHM system for the long-term monitoring of civil infrastructures.

Existing data interpretation methods are limited in their capacity at removing external effects such as traffic and temperature fluctuations. Therefore, it is difficult to distinguish a change in the behavior of a structure from an unusually heavy traffic load, or a seasonal change in temperature. Moreover, they often rely on a threshold mechanism for detecting anomalies; A decision maker triggers an alarm, if the difference between predicted structural responses and observations exceeds predefined threshold values. The lack of capacity to remove external effects when using threshold-based anomaly detection methods typically results in a large

number of false alarms that undermines the viability of the analyses derived from the data used for the continuous monitoring of structures. In addition, even if the data interpretation is carried out as expected, existing anomaly detection methods are not able to quantify probabilistically the capacity of sensors at detecting anomalies with respect to an anomaly's magnitude, its duration, as well as the time required to detect an anomaly, i.e., *quantifying the anomaly detectability*. Moreover, existing methods are often not able to associate the occurrence of an anomaly's magnitude, if detected, to the severity of structural damages, i.e., *quantifying the structural damage detectability*, in order to evaluate the viability and usefulness of a SHM system. Besides, an anomaly's magnitude could be associated to different types of structural damages, yet these methods lack a mechanism to evaluate the performance of a sensor configuration in discriminating between these damages, i.e., *structural damage distinguishability*.

1.1 Objectives and Contributions

This thesis aims at developing data-driven methods that are well suited for monitoring the condition of civil infrastructures based on quantifying two criteria: (i) The anomaly detectability and (ii) The structural damage distinguishability. The former enables minimizing the false detection of anomalies, while maximizing their true detections. The latter allows associating the magnitudes of detected anomalies to the severity levels of structural damages by using multiple sensor configurations. The aforementioned criteria are the basis for the contributions made in this research work. They form the main objectives of this thesis that are summarized as:

- Formulate a methodology for identifying alarm triggering policies in order to detect anomalies. The methodology needs to limit the false alarms, while maintaining the anomaly detectability for anomalies with low magnitudes and long durations.

- Develop a method for characterizing probabilistically the anomaly detectability of a SHM system. The characterization needs to be presented probabilistically as a function of an anomaly's magnitude, its duration, and the time required to detect an anomaly.

- Develop a methodology in order to associate the magnitudes of anomalies to the severity levels of structural damages in order to discriminate the damages that are detectable from the ones that are not.

- Develop a metric in order to quantify the distinguishability of different structural damages over different sensor configurations.

The general flow of this thesis is illustrated in Figure 1.1. The arrows indicate the input-output of eight distinct elements representing the aforementioned objectives. These elements are categorized in four general steps, namely *measurement*, *state inference*, *decision making*, and *damage distinguishability*. The following subsections present the detailed descriptions of these elements.

Measurement

The goal of this step is to present the Jacques Cartier Bridge as well as the measurements collected from three instrumented spans. First, we present the case study because each chapter in this thesis presents methodologies to address particular objectives, followed by a validation for the Jacques Cartier Bridge. Moreover, this step presents the preprocessing procedures for the data collected, identify the possible issues regarding the data acquisition, and discuss the reliability of the collected data for the long-term monitoring of the Jacques Cartier Bridge. The first three elements, from left to right, shown in Figure 1.1, corresponding to this step are: (i) The sensor technology used on the bridge, including the sensors types and their locations, (ii) The Jacques Cartier Bridge involving its general layout and the characteristics of the instrumented spans along with the sensors configuration, and (iii) The data collected on three instrumented spans along with the preprocessing procedure.

State Inference

The goal of this step is to interpret the preprocessed data in order to remove the external effects from the measurements collected on a structure. The proposed interpretation technique uses Bayesian dynamic linear models (BDLM) from the field of machine learning to create empirical models. An empirical model learns to reproduce the behavior of a system, e.g., a bridge, from the measurements recorded on it. In contrast to physics-based models, e.g., finite element model, the empirical models do not contain information about the geometry and a structure's properties. Therefore, empirical models allow resources to be concentrated on the development of a single generic model that can then be used to interpret data from any bridge. As shown in Figure 1.1, the elements corresponding to this step are as follows: (iv) Bayesian dynamic linear model in order to build generic empirical models, and (v) Measurements decomposition consisting in the underlying structural responses and the external effects.

(i)   (ii)   (iii)   (iv)   (v)   (vi)   (vii)   (viii)
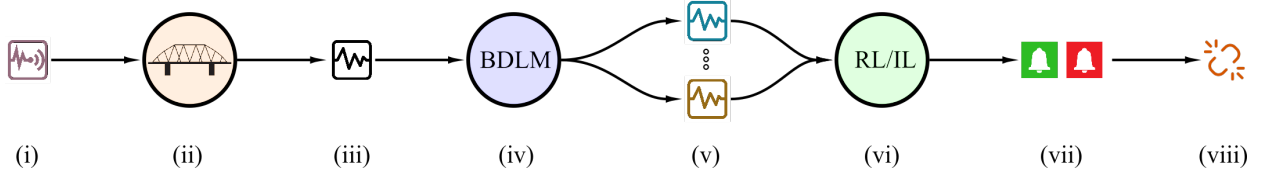
Figure 1.1 The general flow of this thesis composed of eight distinct elements, namely (i) sensor technology, (ii) structural characteristics and sensors configuration, (iii) measurements, (iv) Bayesian dynamic liner models (BDLM), (v) measurement decomposition, (vi) reinforcement learning (RL) or limitation learning (IL), (vii) alarm triggering policies, and (viii) damage distinguishability.

Decision Making

The main objective of this step is to quantify the anomaly detectability of each sensor using the underlying structural responses obtained from the previous step. To this end, the first step is to develop an anomaly detection method capable of operating for a wide range of anomalies. The proposed anomaly detection method uses the reinforcement learning (RL) and imitation learning (IL) approaches from the field of machine learning. In the reinforcement learning approach, instead of defining rule-based alarm-triggering thresholds, we let agents figure out how to optimally trigger alarms while satisfying our preference constraints encoded in a reward function for true anomaly detections, false alarms, and missed alarms. On the other hand, in the imitation learning approach, an agent learns the optimal policy by imitating a demonstration set for which the corresponding optimal policy is issued beforehand. This approach is analogous to a classification problem where the objective is to determine the alarm triggering policy, i.e., to trigger an alarm or not. With the anomaly detection method developed, the detectability of anomalies is quantified through an annual false alarm ratio and a probability of true detections that is a function of the anomaly's magnitude, its duration and the time required to detect an anomaly. The former expresses the number of false alarms that the anomaly detection method yields, and the latter is its capacity at rightfully detecting anomalies. Upon addressing the objectives of this section, the anomaly detectability quantification will serve for the next step. From Figure 1.1, the associated elements are: (vi) Reinforcement learning or imitation learning approaches as the basis for detecting anomalies, and (vii) Alarm triggering policies which dictate the optimal action, in this case triggering an alarm in the presence of an anomaly or not triggering an alarm otherwise.

Damage Distinguishability

The ensemble of methods developed in the previous steps enables associating the magnitudes of detected anomalies to the severity levels of structural damage scenarios, and evaluating the capacity of different sensor configurations in discriminating between these damages. Therefore, we are able to judge the performance of various sensor types along with their placements in order to maximize the distinguishability between different structural damage scenarios; It enables discriminating between the damage scenarios that will be detectable from those that will not, thus setting realistic expectations with respect to the capacity of a structural health monitoring system. In Figure 1.1, this step is associated with the last element: (viii) damage distinguishability.

1.2 Thesis Structure

In this thesis, first we propose an anomaly detection method for interpreting data collected on civil infrastructures. In addition, we present methodologies in order to quantify the anomaly detectability of sensors with respect to two metrics, namely the *annual false alarm ratio* and the *probability of true anomaly detections.* The former quantifies the number of anomalies per year that are wrongfully detected. The latter quantifies the probability of rightfully detecting an anomaly with respect to an anomaly's magnitude, its duration, and the time required to detect an anomaly. This thesis validates the methodologies developed on three instrumented spans of the Jacques Cartier Bridge in Canada. Upon quantifying the anomaly detectability, we present methodologies in order to associate the magnitudes of detected anomalies to the severity levels of structural damages, and maximize the distinguishability between them using multiple sensor configurations. Figure 1.2 illustrates the structure of the current thesis with respect to the objectives presented in Section 1.1. The four aforementioned steps are shown along with their corresponding chapters. The arrows show the flow of theses steps. Chapter 2 provides the literature review for the existing methods along with the necessary background for Chapters 3-6. Chapter 3 presents the Jacques Cartier Bridge instrumentation as well as the acquired measurements. Furthermore, this chapter presents the analysis carried out on the time series data obtained from the sensors. Chapter 4 presents the methodologies developed based on the reinforcement learning approach for detecting anomalies as well as for the anomaly detectability quantification. Chapter 5 presents the methodologies developed based on the imitation learning approach for detecting anomalies, and for the anomaly detectability quantification. Chapter 6 presents the damage scenario detectability and distinguishability methodologies based on the anomaly detection methods developed in the previous chapters. Chapter 7 presents the conclusion of the current research
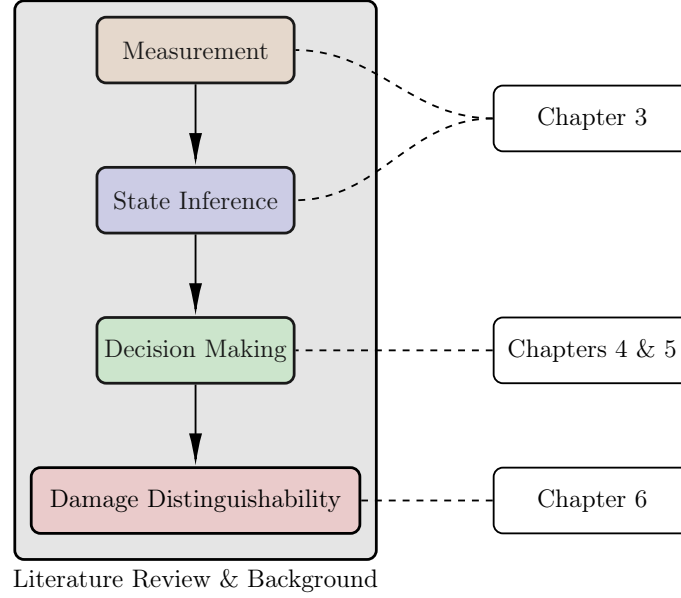
Figure 1.2 The structure of the current report. The project's objectives are divided in four steps, namely measurement, state inference, decision making, and damage distinguishability.

project along with its limitations and future work.

Co-Authored Papers

A part of the work presented in this thesis has already been published. The list of the published co-authored papers are

- S. Khazaeli and J.-A. Goulet, "Damage detection for structural health monitoring using reinforcement and imitation learning," *Engineering application of artificial intelligence*, 2022 (submitted).

- S. Khazaeli, L. H. Nguyen, and J.-A. Goulet, "Anomaly detection using state-space models and reinforcement learning," *Structural Control and Health Monitoring*, vol. 28, no. 6, p. e2720, 2021.

- L. H. Nguyen, I. Gaudot, S. Khazaeli, and J.-A. Goulet, "A kernel-based method for modeling non-harmonic periodic phenomena in Bayesian dynamic linear models," *Frontiers in Built Environment*, vol. 5, p. 8, 2019.

CHAPTER 2   Literature Review and Background

This chapter presents a review of existing structural health monitoring strategies along with the methods for detecting anomalies and their limitations. In addition, this chapter presents the limitations of existing methods for detecting structural damages. Furthermore, this chapter provides the background for the methodologies developed in this thesis. In this regard, it presents the fundamentals of Bayesian dynamic linear models (BDLM) that are empirical models based on the data collected on a structure in order to separate the external effects such as the temperature from the structural responses. In addition, it presents the limitations of a BDLM for detecting anomalies. Afterwards, This chapter presents the fundamentals associated with the reinforcement learning (RL) and imitation learning (IL) approaches, and discuss how these approaches may overcome the existing limitations corresponding to detecting anomalies.

2.1   Structural Health Monitoring

The deterioration and degradation of civil infrastructures such as bridges are responsible for important societal expenses. According to the latest Canadian infrastructure report card (CIRC) [14], nearly 40% of roads and bridges are in fair or worse condition. An independent report carried out by Trisura [15] has indicated that according to the Canadian federation of municipalities (FCM), there is a need for more than $250 billion dollars of investment in order to bring Canadian infrastructures to an acceptable condition. Moreover, Risk Analytica [16] has claimed that the Canadian government needs to invest annually 5.1% of its gross domestic product (GDP) on infrastructure over a span of 50 years in order to reach a maximum GDP growth, from which 22% of this investment should be allocated for repairs and maintenance.

In the U.S., the American road and transportation builder association (ARTBA) [17] has found that more than 46000 bridges are in poor condition, for which one out of three needs to be repaired or replaced. The American society of civil engineers (ASCE) estimated the gap in investment to maintain infrastructures in a good condition to be 2.5 trillion dollars over the next 10 years [18]. The latest ASCE estimations have revealed that underinvestment in infrastructures will cost more than 10 trillion dollars by 2039. Despite the need for such a large amount of investments, limited financial resources prohibit the maintenance of all infrastructures at the same time. To address this limitation, an approach is to *monitor* infrastructures in order to assess their condition and enable the data-driven allocation of resources.

Structural health monitoring (SHM) aims at providing information about structural condition for decision makers in order to support the operation, maintenance, and the replacement of structures [10, 13]. SHM uses sensors to collect measurements about the structural responses and their surrounding environment. The measurements are then interpreted to infer the condition of a structure over time. In this regard, an *anomaly*, is defined as a change in the structural responses from a *reference* condition that is referred to as the *normal condition.* Changes from the normal condition of a structure are often associated with various damages that can developed over a period spanning several years. Therefore, a SHM system should focus on the long-term monitoring of structures in order to detect anomalies at an early stage before they become critical. In the context of anomaly detection, a true detection consists in correctly identifying an anomaly when it occurs. In contrast, a false detection, i.e., a *false alarm*, occurs when a decision maker wrongfully identifies the presence of an anomaly. Therefore, detecting anomalies and avoiding false alarms is the primary objective when developing anomaly detection methods in the context of structural health monitoring.

A SHM system requires building a model in order to analyze structural responses; From the modeling perspective, two primary SHM approaches are identified in the literature, namely *physics-based* and *data-driven* models [13, 19]. Physics-based models rely on physical principles to build a model for a structure. Such a model is used to analyze structural responses in order to gain knowledge about *damage sensitive properties* associated with the structure's mass, stiffness, and damping. The finite element (FE) method [20] is among popular approaches to create physics-based models for structures. Once a FE model is built, structural responses obtained from sensors are used to update damage sensitive properties of the model, so they coincide with the measured responses from the sensors [21, 22]. Afterwards, a decision maker can use changes in damage sensitive properties during the updating process in order to identify the presence of damages and their potential locations. However, physics-based models have a limited ability to represent structural responses over time when subject to complex environmental conditions such as traffic load and air temperature, which leads to a limited scalability [19]. In addition, physics-based models are often case-specific in the sense that a decision maker needs to treat each structure individually and build a separate model for each of them. This results in additional complexities and costs during the model building and re-calibration steps.

In contrast, data-driven models solely rely on structural responses without the necessity of building a physics-based model. Many methods from the field of statistics and machine learning have been developed in order to detect anomalies in time series data. The reader is referred to [19, 23, 24] for a detailed review of these methods. The vast majority of existing data-driven anomaly detection approaches are based on a threshold-based alarm-triggering

paradigm. In general, these methods create a model from a training dataset representing the normal signature of a structure. Training datasets often consist in *damage sensitive features* extracted from dynamic responses of structures. For instance, Balsamo et al. [10] used Mel-frequency cepstral coefficients extracted from structural vibration responses as damage sensitive features. Such coefficients are defined in the context of cepstrum domain from the field of speech recognition, and have been showed to be effective in the context of structural health monitoring applications. Once a model is built from training datasets, the newly collected data are analyzed and tested against the model in order to identify whether or not the predictions fit the test data within an acceptable threshold. For instance, Morgantini et al. [25] and Achilli et al. [26] used the squared Mahalanobis distance as metric to measure the distance between the model predictions and the test dataset, and the assessment regarding the presence of an anomaly is carried out by comparing this distance with a predefined threshold. The reader is referred to [19, 27–31] for examples of threshold-based anomaly detection methods. The key limitation of these methods is not related to the choice of a particular mathematical formulation for modeling time series data, but to the way anomalies are detected; The presence or absence of an anomaly is inferred based on the difference between the predicted values returned by the model and those observed on a structure. This approach is subject to a tradeoff between the frequency of false alarms and the anomaly detectability. This aspect is the key factor limiting the scalability of a SHM system for anomaly detection as outliers and changes in the conditions surrounding a structure cause frequent false alarms that need to be mitigated by raising the detection threshold, which in turn reduces the detectability of anomalies. One method that has tackled several of the limitations with respect to the threshold-based approaches for anomaly detection relies on the Bayesian dynamic linear model (BDLM) [32] that will be reviewed in Section 2.2.

Another limitation of existing data-driven methods is that when a model is built, it cannot be continuously adapted as new data are collected [19, 33–35]. In addition, they need labeled data, where anomalies have been identified, which are typically not available for many civil structures or costly to obtain. In particular, anomalies in infrastructures are rare events, and yet in many cases, existing methods require labels for the structural conditions resulting in a limited scalability [10, 30, 33, 36, 37].

Another limitation is that most existing methods have not been developed or validated with real operational conditions, where a full-scale structure is studied in its operational environment, and where it is subject to loading such as traffic, and external effects such as temperature and weather changes. Moreover, existing methods typically study ad-hoc punctual and sudden damage scenarios whereas in practice we are interested in detecting anomalies that have a small magnitude, which develop over a time-span of years. For examples, Farrar and

Worden [19] have employed statistical pattern recognition to find the anomalies and correlated them to different damage scenarios on a benchmark laboratory structure in the Los Alamos National Laboratory, USA. Ghazi and Büyüköztürk [28] have employed hypothesis-testing for damage detection of a laboratory structure using energy-based non-linear features. Balsamo et al. [10] and Li et al. [31] have employed cepstural damage sensitive features and examined them on a simulated and a laboratory shear-type building. Overall, there is a lack of methods that been validated in real-life operational condition for a wide range of anomaly scenarios.

A common approach to quantify the performance of anomaly detection methods is to use metrics such as accuracy, F1-score, and recall, in order to quantify true and false anomaly detections [38–40]. In general, an effective anomaly detection method results in metrics indicating a high early true detection rate, while keeping a low false alarm rate. However, existing metrics do not account for anomaly characteristics such as the anomaly's *magnitude*, its *duration* and *time to detection*; Anomalies affect the measurements collected on a structure by creating a deviation in the behavior of structures representing the normal condition. This deviation can be quantified in terms of its magnitude, which is referred to as the *anomaly magnitude*. In addition, such a deviation occurs over a period of time known as the *anomaly duration*. The *time to detection* refers to the time that a SHM method takes in order to identify the presence of an anomaly after its starting time.

A structural health monitoring system should select sensor configurations such that they (i) lower the costs associated with the type and number of sensors, (ii) minimize false detections of anomalies, while maximize their true detections, (iii) discriminate between different structural damages, and (iv) separate the environmental effects from the structural measurements [11]. Therefore, data interpretation and anomaly detection methods proposed in the literature play the key role in order to determine the best sensor configuration. One of the most popular methods in detecting anomalies is associated with vibration-based monitoring [41] for which the idea is that any mechanical change in the structure is reflected in the changes in its dynamic characteristics such as mode shapes. A mode shape refers to the deformation of a structure vibrating at its natural frequency [42]. When a damage occurs, the natural frequency changes, hence the corresponding mode shape. Kammer [43] employed a metric called the *Fisher information matrix* (FIM) in order to track the changes in the structural mode shapes. A change in the FIM metric is an indication of a change in the mode shapes so as in the dynamic characteristics of a structure. Therefore, changes in a FIM metric are implicitly associated with anomalies that occurred on a structure. As a result, the best sensor configuration is the one which results in a higher FIM because it maximizes the information corresponding to the data acquired from the locations yielding

best modal identification [44]. Another metric employed for identifying an optimal sensor configuration relies on the signal-to-noise ratio (SNR) [45]. This metric is obtained by modeling the measurements at different sensor locations as a function of the spatial location of the sensors. A higher value for the SNR indicates that the location of the sensor results in gaining more information about the structural condition. Using FIM and SNR metrics in order to identify the best sensor configuration is an optimization problem for which the objective is to maximize the values of these metrics. The reader is referred to [11, 45–48] for a detailed explanation about formulating these methods and solving the resulting optimization problem. However, using such metrics only paints a part of the picture as there is only an indirect relationship between these metrics and the anomaly detectability; A decision maker is not informed about the presence of an anomaly by only using the SNR metric, and the FIM metric is not capable of effectively discriminating between different structural damages. In addition, structural damages occur locally in most cases and therefore, a change in the local dynamic properties of a structure may not be detected by tracking the changes in the FIM metric, and consequently changes in structural mode shapes. Moreover, these metrics do not necessary carry information regarding the severity levels of the structural damages that occurred on a structure.

In general, existing approaches lack a systematic way for probabilistically characterizing the anomaly detectability as a function of its characteristics such as the anomaly's magnitude, its duration and the time to detection, and further associate the magnitude of an anomaly to the severity levels of structural damages. The remaining sections of this chapter provide the theoretical backgrounds for the methodologies developed in this thesis.

2.2 Bayesian Dynamic Linear Models

In order to detect anomalies from SHM data, Goulet [32] has employed a particular type of state-space models called Bayesian dynamic linear models (BDLMs) which assume that all variables follow *multivariate Gaussian distributions* and all the equations are *linear*. The core idea behind BDLMs is to use the knowledge of a time-series kinematics in order to predict the evolution of its sub-components through time and combine this information with the information contained in observations. Using sub-components allows decomposing structural responses into the superposition of *reversible* and *irreversible* behaviors. Reversible behaviors are typically due to external effects such as temperature and loading, whereas irreversible ones are the result of changes in a structure itself. Irreversible responses are decomposed using generic baseline components including a baseline, a baseline's rate of change, and a baseline's acceleration components [32, 49, 50]. They represent the structure's degradation,

degradation's speed, and degradation's acceleration, respectively. Reversible responses are decomposed using either a periodic [49], a dynamic regression [51], or a kernel regression [52] component. In addition, in a BDLM, time-dependent model errors are handled by an autoregressive component. BDLMs are able to consider more complex model structures than methods such as autoregressive models (ARMs) and autoregressive moving average (ARMA) models [32, 38]. In fact, these models can be seen as special cases of BDLMs for which only an autoregressive component is used in building a dynamic model.

2.2.1 Mathematical Formulation

BDLMs use generic sub-components for representing the underlying behaviors for the observation vector $\mathbf{y}_t = [y_1 \ y_2 \ \cdots \ y_\mathtt{Y}]^\mathsf{T}_t$ that consists in $\mathtt{Y}$ observations at a time t. Sub-components are assembled in order to form the *hidden state* vector $\mathbf{x}_t = [x_1 \ x_2 \ \cdots \ x_\mathtt{X}]^\mathsf{T}_t$ consisting in $\mathtt{X}$ hidden states. The term *hidden* refers to the fact that the corresponding variables are not observed directly. At a time t, the estimation is based on the information coming from two sources: A linear dynamic model called a *transition model* that predicts the evolution of the hidden states variables across time, and a linear *observation model* based on the empirical observations $\mathbf{y}_t$ in order to update the knowledge of the hidden state variables. The transition model is defined following

$$\mathbf{x}_t = \mathbf{A}_t\mathbf{x}_{t-1} + \mathbf{w}_t, \ \ \mathbf{w} : \boldsymbol{W} \sim \mathcal{N}(\mathbf{w}; \mathbf{0}, \mathbf{Q}_t), \tag{2.1}$$

where $\mathbf{A}_t$ is the transition matrix and $\mathbf{w}_t$ is the transition model error vector following a Gaussian distribution with zero mean and covariance matrix $\mathbf{Q}_t$. The observation model follows

$$\mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{v}_t, \ \ \mathbf{v} : \boldsymbol{V} \sim \mathcal{N}(\mathbf{v}; \mathbf{0}, \mathbf{R}_t), \tag{2.2}$$

in which $\mathbf{C}_t$ is the observation matrix and $\mathbf{v}_t$ is a vector of observation errors following a Gaussian distribution with zero mean and covariance matrix $\mathbf{R}_t$. The formulation of the model matrices $\mathcal{M} = \{\mathbf{A}_t, \mathbf{C}_t, \mathbf{Q}_t, \mathbf{R}_t\}$ is based on the assembly of generic sub-components which can model behaviors such as a structure's degradation, a degradation's speed, a degradation's acceleration, periodic phenomena, and so on. The reader is referred to [32, 35, 49, 50] for a complete review of the different sub-components available.

Consider Figure 2.1 where the observed structural responses y_t are shown by gray circles. The units of the observations depend on the sensor type, e.g., mm for elongation or °d for inclination. Figures 2.1b-2.1d illustrate the baseline (B) generic sub-components including the structure's degradation $x_t^\mathtt{B}$, degradation's speed $\dot{x}_t^\mathtt{B}$, and degradation's acceleration $\ddot{x}_t^\mathtt{B}$.

(a) observations

(b) degradation

(c) degradation's speed

(d) degradation's acceleration
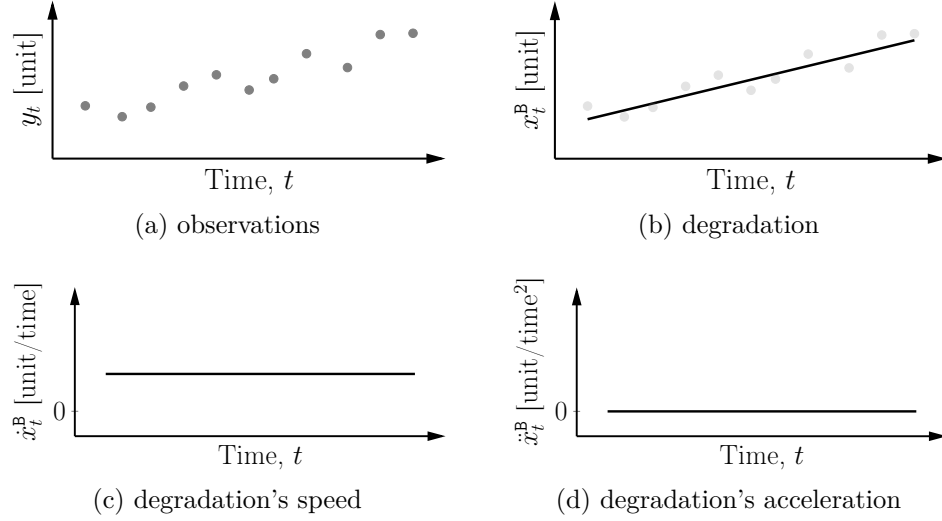
Figure 2.1 Illustration of the kinematics of a dynamic model including: (a) observations, (b) structure's degradation (c) structure's degradation's speed, and (d) structure's degradation's acceleration.

The dot notation indicates derivatives with respect to the time such that $\dot{x}_t^{\text{B}} = \frac{d}{dt}x_t^{\text{B}}$ and $\ddot{x}_t^{\text{B}} = \frac{d^2}{dt^2}x_t^{\text{B}}$. In order to simplify the illustration, the uncertainty around the estimations is not shown in Figures 2.1b-2.1d.

The structure's degradation, its speed, and its acceleration are hidden state variables that jointly define the *degradation's kinematics* of the system. They represent the transition model describing the evolution of the hidden state variables vector $\mathbf{x}_t = [x_t^{\text{B}} \ \dot{x}_t^{\text{B}} \ \ddot{x}_t^{\text{B}}]^{\mathsf{T}}$. The general form of such kinematics in a discrete-time domain with a time step length Δt follows

$$\text{Kinematics}: \begin{cases} x_t^{\text{B}} = x_{t-1}^{\text{B}} + \dot{x}_{t-1}^{\text{B}}\Delta t + \ddot{x}_{t-1}^{\text{B}}\dfrac{\Delta t^2}{2} + w_t^{\text{B}} \\[2mm] \dot{x}_t^{\text{B}} = \dot{x}_{t-1}^{\text{B}} + \ddot{x}_{t-1}^{\text{B}}\Delta t + w_t^{\dot{\text{B}}} \\[2mm] \ddot{x}_t^{\text{B}} = \ddot{x}_{t-1}^{\text{B}} + w_t^{\ddot{\text{B}}} \end{cases}, \quad \forall t. \tag{2.3}$$

In Equation 2.3, w denotes the transitions errors; They correspond to three hidden state variables namely structure's degradation error w_t^{B}, the degradation's speed error $w_t^{\dot{\text{B}}}$, and the degradation's acceleration error $w_t^{\ddot{\text{B}}}$. It is common to write Equation 2.3 in the matrix form

as

$$\text{Kinematics}: \underbrace{\begin{bmatrix} x_t^{\text{B}} \\ \dot{x}_t^{\text{B}} \\ \ddot{x}_t^{\text{B}} \end{bmatrix}}_{\mathbf{x}_t} = \underbrace{\begin{bmatrix} 1 & \Delta t & \frac{\Delta t^2}{2} \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{A}_t} \underbrace{\begin{bmatrix} x_{t-1}^{\text{B}} \\ \dot{x}_{t-1}^{\text{B}} \\ \ddot{x}_{t-1}^{\text{B}} \end{bmatrix}}_{\mathbf{x}_{t-1}} + \underbrace{\begin{bmatrix} w_t^{\text{B}} \\ w_t^{\dot{\text{B}}} \\ w_t^{\ddot{\text{B}}} \end{bmatrix}}_{\mathbf{w}_t}, \tag{2.4}$$

where, $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$ are respectively the hidden state vectors at time t and $t-1$; $\mathbf{A}_t$ is the transition matrix and $\mathbf{w}_t$ is the error vector at time t. Moreover, the observations y_t are decomposed into its underlying behavior and error terms following the observation model

$$y_t = \underbrace{\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}}_{\mathbf{C}_t} \begin{bmatrix} x_t^{\text{B}} \\ \dot{x}_t^{\text{B}} \\ \ddot{x}_t^{\text{B}} \end{bmatrix} + v_t, \tag{2.5}$$

where, $\mathbf{C}_t$ is the observation matrix and v_t is the observation error. Figure 2.1 is an example of a *trend-stationary regime* with respect to the degradation's kinematics. The kinematics of the hidden state variables for a stationary regime follows a constant degradation's speed for all time steps so that the degradation's acceleration $\ddot{x}_t^{\text{B}} = w_t^{\ddot{\text{B}}} = 0$ over time. Note that, in this thesis, both terms *stationary* and *trend-stationary* are used interchangeably. The transition matrix and error vector presented in Equation 2.4 are re-written as

$$\mathbf{A}_t = \begin{bmatrix} 1 & \Delta t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ \mathbf{w}_t = \begin{bmatrix} w_t^{\text{B}} \\ w_t^{\dot{\text{B}}} \\ 0 \end{bmatrix}. \tag{2.6}$$

In order to keep the hidden state estimation procedure simple, it is helpful to rely on a shorthand notation for the value that a variable can take from time 1 to time t by a subscript $1{:}t$. Therefore, $y_{1:t} \equiv \{y_1, y_2, \cdots, y_t\}$ defines a set containing t observations. $\boldsymbol{\mu}_{t|t} \equiv \mathbb{E}[\boldsymbol{X}_t | y_{1:t}]$ defines the expected value of the hidden states vector at time t, conditional on all the data ranging from time 1 up to time t. Note that in the case where we are interested in times from 1 up to $t-1$, the notation changes for $\boldsymbol{\mu}_{t|t-1} \equiv \mathbb{E}[\boldsymbol{X}_t | y_{1:t-1}]$. Analogously, $\boldsymbol{\Sigma}_{t|t} \equiv \text{cov}[\boldsymbol{X}_t | y_{1:t}]$ denotes the covariance of a hidden states vector at time t, conditional on all available observations starting from time 1 up to time t. The hidden state estimation is a recursive process in which the prior knowledge at time t is obtained from the posterior knowledge at time $t-1$. This task is carried on using the *Kalman filter* (KF) [53], which is an iterative two-step mathematical process consisting in the *prediction* and *update* steps.

The prediction step computes the joint prior knowledge $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ at time t following

$$
\begin{aligned}
p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) &= \mathcal{N}(\mathbf{x}_t; \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}) \quad &\text{Prior state estimate} \\
\boldsymbol{\mu}_{t|t-1} &= \mathbf{A}_t\boldsymbol{\mu}_{t-1|t-1} \quad &\text{Prior expected value} \\
\boldsymbol{\Sigma}_{t|t-1} &= \mathbf{A}_t\boldsymbol{\Sigma}_{t-1|t-1}\mathbf{A}_t^\intercal + \mathbf{Q}_t \quad &\text{Prior covariance.}
\end{aligned}
\tag{2.7}
$$

The update step evaluate the posterior knowledge $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ at time t following

$$
\begin{aligned}
p(\mathbf{x}_t|\mathbf{y}_{1:t}) &= \mathcal{N}(\mathbf{x}_t; \boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t}) \quad &\text{Posterior state estimate} \\
\boldsymbol{\mu}_{t|t} &= \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t\mathbf{r}_t \quad &\text{Posterior expected value} \\
\boldsymbol{\Sigma}_{t|t} &= (\mathbf{I} - \mathbf{K}_t\mathbf{C}_t)\boldsymbol{\Sigma}_{t|t-1} \quad &\text{Posterior covariance} \\
\mathbf{r}_t &= \mathbf{y}_t - \hat{\mathbf{y}}_t \quad &\text{Innovation vector} \\
\hat{\mathbf{y}}_t &= \mathbb{E}[\mathbf{y}_t|\mathbf{y}_{1:t-1}] = \mathbf{C}_t\boldsymbol{\mu}_{t|t-1} \quad &\text{Predicted observations vector} \\
\mathbf{K}_t &= \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}_t^\intercal\mathbf{G}_t^{-1} \quad &\text{Kalman gain matrix} \\
\mathbf{G}_t &= \mathbf{C}_t\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}_t^\intercal + \mathbf{R}_t \quad &\text{Innovation covariance matrix.}
\end{aligned}
\tag{2.8}
$$

The hidden state estimation can be summarized by

$$
(\boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t}, \mathcal{L}_t) = \text{Filter}(\boldsymbol{\mu}_{t-1|t-1}, \boldsymbol{\Sigma}_{t-1|t-1}, \mathbf{y}_t, \mathbf{A}_t, \mathbf{C}_t, \mathbf{Q}_t, \mathbf{R}_t),
\tag{2.9}
$$

where $\mathcal{L}_t$ is the marginal likelihood that will be further described in Section 2.2.3. A key aspect of BDLMs is their ability to detect anomalies without relying on detection thresholds; Instead, it relies on a *regime switching* approach [2], where several models displaying different degradation kinematics are evaluated in parallel, and the probability of each of these models can be estimated in real-time as data become available.

2.2.2 Regime Switching

Figure 2.2 shows an example of two regimes: a *trend-stationary* regime similar to the one shown Figure 2.1, where the degradation's speed remains constant, and a *non-stationary regime*, where the degradation's speed varies over time. The non-stationary regime is shown by the red dashed lines in Figures 2.2b–2.2d. Note that the transition matrix $\mathbf{A}_t$ and error vector $\mathbf{w}_t$ for the stationary and non-stationary regimes follow the ones presented in Equations 2.4 and 2.6, respectively. In Figure 2.2, the dynamic model starts with a constant degradation's speed during the stationary regime and then switches to a non-stationary regime with a locally constant degradation's acceleration before it returns to the stationary regime
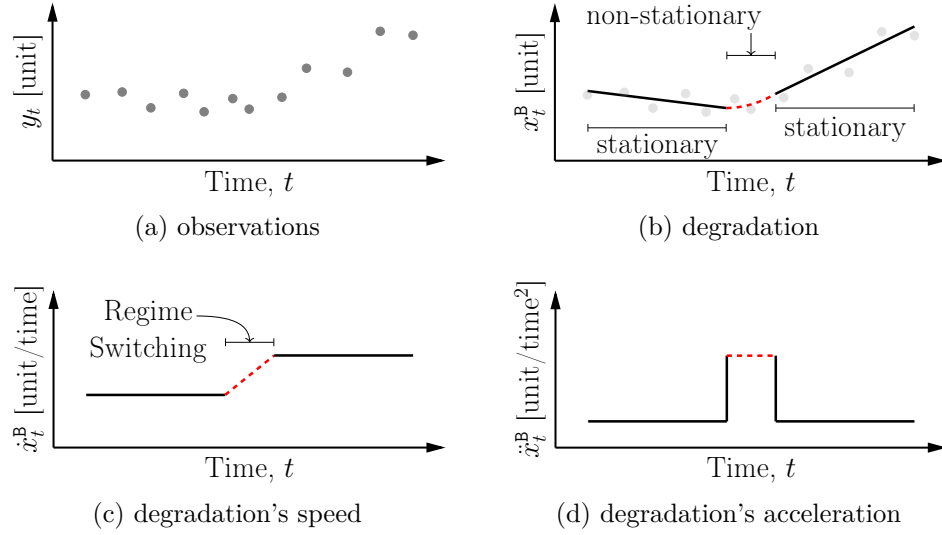
Figure 2.2 Illustration of the kinematics of two dynamic models including: (a) observation, (b) structure's degradation (c) structure's degradation's speed, and (d) structure's degradation's acceleration.

with a different constant speed. In BDLMs, two dynamic models are built corresponding to the two regimes for which the *probabilities* are evaluated using the switching Kalman filter (SKF) [2].

Figure 2.3 shows an idealized representation of the regime switching probabilities associated with Figure 2.1; the probability of the stationary regime is one during all time steps, except for the central portion, where the model switches to the non-stationary regime. In this configuration, the task of *anomaly detection* consists in identifying the presence of a switch from a stationary regime to a non-stationary one. Figure 2.3 shows the time of detection $t = j$ when the increase in probability for the non-stationary regime is used to trigger an
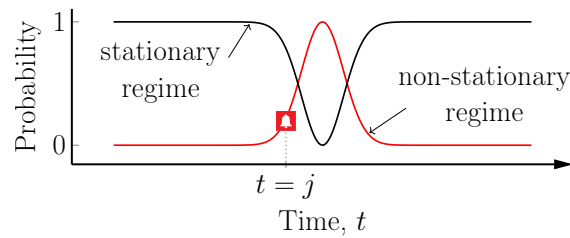
Figure 2.3 Illustration of the probability of regime switching probabilities along with the time $t = j$ when the regime switch is detected.

alarm. The remaining part of this section presents the formalism of the switching Kalman filter, and discusses its current limitations in the context of anomaly detection.

The SKF models non-stationary system responses by estimating at each time t, the probability of each regime along with their hidden state values. Each regime is defined by its own *model matrices.* Consider Figure 2.2b where a system that has two possible regimes are identified by the regime state variable $s_t \in \{1, 2\}$, where 1: stationary ($\mathbf{s}$) or 2: non-stationary ($\mathbf{ns}$). Each regime state s_t has a probability $\pi_t^{s_t} \equiv p(s_t|\mathbf{y}_{1:t})$. Note that the relationship between the probability of the stationary and non-stationary regimes is $\pi_t^1 = 1 - \pi_t^2$ indicating that the two possible regime states are mutually exclusive and collectively exhaustive. Therefore, at a time t, the objective of the SKF is to estimate: (i) the hidden state variables vector $\mathbf{x}_t$ and (ii) the regime state probability $\pi_t^{s_t}$ associated with each regime state s_t. Analogously, the hidden state estimation in SKF consists in a two-step mathematical process; The first step computes the joint prior knowledge for both regimes $p\left(\mathbf{x}_t^{i(j)}|\mathbf{y}_{1:t-1}\right)$ at time t following

$$
\begin{aligned}
p\left(\mathbf{x}_t^{i(j)}|\mathbf{y}_{1:t-1}\right) &= \mathcal{N}\left(\mathbf{x}_t^{i(j)}; \boldsymbol{\mu}_{t|t-1}^{i(j)}, \boldsymbol{\Sigma}_{t|t-1}^{i(j)}\right) && \text{Prior state estimate} \\
\boldsymbol{\mu}_{t|t-1}^{i(j)} &= \mathbf{A}_t^j \boldsymbol{\mu}_{t-1|t-1}^i && \text{Prior expected value} && (2.10) \\
\boldsymbol{\Sigma}_{t|t-1}^{i(j)} &= \mathbf{A}_t^j \boldsymbol{\Sigma}_{t-1|t-1}^i \left(\mathbf{A}_t^j\right)^{\mathsf{T}} + \mathbf{Q}_t^{i(j)} && \text{Prior covariance,}
\end{aligned}
$$

where, the superscript inside the parentheses $^{i(j)}$ indicates the current state $s_t = j$ at time t given the state $s_{t-1} = i$ at time $t-1$. The update step evaluates the posterior knowledge for both regimes $p\left(\mathbf{x}_t^{i(j)}|\mathbf{y}_{1:t}\right)$ at time t following

$$
\begin{aligned}
p\left(\mathbf{x}_t^{i(j)}|\mathbf{y}_{1:t}\right) &= \mathcal{N}\left(\mathbf{x}_{t|t}^{i(j)}; \boldsymbol{\mu}_{t|t}^{i(j)}, \boldsymbol{\Sigma}_{t|t}^{i(j)}\right) && \text{Posterior state estimate} \\
\boldsymbol{\mu}_{t|t}^{i(j)} &= \boldsymbol{\mu}_{t|t-1}^{i(j)} + \mathbf{K}_t^{i(j)} \mathbf{r}_t^{i(j)} && \text{Posterior expected value} \\
\boldsymbol{\Sigma}_{t|t}^{i(j)} &= \left(\mathbf{I} - \mathbf{K}_t^{i(j)} \mathbf{C}_t^j\right) \boldsymbol{\Sigma}_{t|t-1}^{i(j)} && \text{Posterior covariance} \\
\mathbf{r}_t^{i(j)} &= \mathbf{y}_t - \hat{\mathbf{y}}_t^{i(j)} && \text{Innovation vector} && (2.11) \\
\hat{\mathbf{y}}_t^{i(j)} &= \mathbb{E}[\mathbf{y}_t|\mathbf{y}_{1:t-1}] = \mathbf{C}_t^j \boldsymbol{\mu}_{t|t-1}^{i(j)} && \text{Predicted observations vector} \\
\mathbf{K}_t^{i(j)} &= \boldsymbol{\Sigma}_{t|t-1}^{i(j)} \left(\mathbf{C}_t^j\right)^{\mathsf{T}} \left(\mathbf{G}_t^{i(j)}\right)^{-1} && \text{Kalman gain matrix} \\
\mathbf{G}_t^{i(j)} &= \mathbf{C}_t^j \boldsymbol{\Sigma}_{t|t-1}^{i(j)} \left(\mathbf{C}_t^j\right)^{\mathsf{T}} + \mathbf{R}_t^j && \text{Innovation covariance matrix.}
\end{aligned}
$$

The SKF hidden state estimation procedure can be summarized by

$$
(\boldsymbol{\mu}_{t|t}^{i(j)}, \boldsymbol{\Sigma}_{t|t}^{i(j)}, \mathcal{L}_t^{i(j)}) = \text{Filter}(\boldsymbol{\mu}_{t-1|t-1}^j, \boldsymbol{\Sigma}_{t-1|t-1}^i, \mathbf{y}_t, \mathbf{A}_t^j, \mathbf{C}_t^j, \mathbf{Q}_t^{i(j)}, \mathbf{R}_t^j), \qquad (2.12)
$$

where, $\mathcal{L}_t^{i(j)}$ is the marginal likelihood. Note that in case of single-regime models, the SKF formulation in Equation 2.12 is reduced to the filtering formulation as presented in Equation 2.9 by setting $i = j = 1$. For the SKF, the transition matrix is defined following

$$
\mathbf{Z} = \begin{bmatrix} z^{i(i)} & z^{i(j)} \\ z^{j(i)} & z^{j(j)} \end{bmatrix}, \quad \mathbf{Z}^{i(j)} = 1 - \mathbf{Z}^{i(i)}, \tag{2.13}
$$

where, $\mathbf{Z}^{i(j)} = p(s_t = j | s_{t-1} = i)$ denotes the prior probability of transitioning from a state i at time $t - 1$ to a state j at time t. The joint regime probability $\mathbf{M}_{t-1,t|t}^{i(j)} = p(s_{t-1} = i, s_t = j | \mathbf{y}_{1:t})$ is evaluated from the combination of the likelihood of observations $\mathcal{L}_{t|t}^{i(j)}$, the transition probabilities $\mathbf{Z}^{i(j)}$, and the prior probability of the states $\pi_{t-1|t-1}^i$ following

$$
\mathbf{M}_{t-1,t|t}^{i(j)} = \frac{\mathcal{L}_{t|t}^{i(j)} \cdot \mathbf{Z}^{i(j)} \cdot \pi_{t-1|t-1}^i}{\underbrace{\sum_i \sum_j \mathcal{L}_{t|t}^{i(j)} \cdot \mathbf{Z}^{i(j)} \cdot \pi_{t-1|t-1}^i}_{\text{Normalization constant}}}. \tag{2.14}
$$

Therefore, the probability of each regime is obtained by marginalizing the joint regime probabilities as

$$
\pi_{t|t}^j = \sum_i \mathbf{M}_{t-1,t|t}^{i(j)}. \tag{2.15}
$$

Figure 2.4a illustrates a limitation of Equation 2.12, where the number of models combinations increases exponentially with the number of time steps. For example, starting from two models at time $t - 1$ the number of models reaches eight after two successive time steps at time $t + 1$. To address this issue, the SKF employs the *collapse* step using the generalized pseudo Bayesian algorithm of order two [54]. Figure 2.4b illustrates the collapse step for two models between time steps $t - 1$ and t; Applying the SKF on two models, yields four models each having a posterior mean vector, covariance matrix, and joint regime probabilities. The collapse step receives them as the input and returns two models each having a posterior mean value vector, covariance matrix, and a probability associated with each regime state. The collapse step can be summarized by

$$
(\boldsymbol{\mu}_{t|t}^j, \boldsymbol{\Sigma}_{t|t}^j, \pi_{t|t}^j) = \text{Collapse}(\boldsymbol{\mu}_{t|t}^{i(j)}, \boldsymbol{\Sigma}_{t|t}^{i(j)}, \mathbf{W}_{t-1|t}^{i(j)}), \tag{2.16}
$$
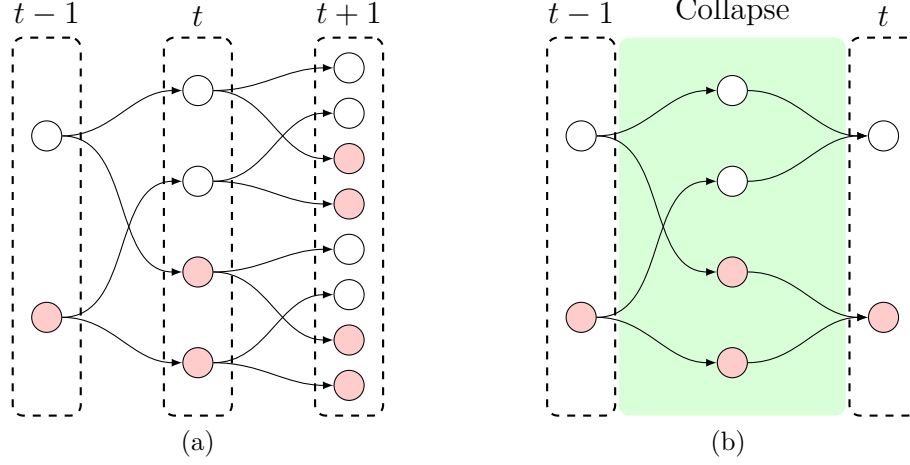
Figure 2.4 Illustration of the number of models: (a) exponential growth of the models within time and (b) Collapse approximation step between time $t-1$ and its successor t. Figure is reproduced from Goulet [1].

where,

$$
\begin{aligned}
\mathbf{W}_{t-1|t}^{i(j)} &= \frac{\mathbf{M}_{t-1,t|t}^{i(j)}}{\pi_{t|t}^{j}} \\
\boldsymbol{\mu}_{t|t}^{j} &= \boldsymbol{\mu}_{t|t}^{i(j)} \cdot \mathbf{W}_{t-1|t}^{i(j)} \qquad \text{Posterior expected value} \\
\mathbf{m} &= \boldsymbol{\mu}_{t|t}^{i(j)} - \boldsymbol{\mu}_{t|t}^{j} \\
\boldsymbol{\Sigma}_{t|t}^{j} &= \sum_{j} \left[\mathbf{W}_{t-1|t}^{i(j)} \cdot (\boldsymbol{\Sigma}_{t|t}^{i(j)} + \mathbf{m}\mathbf{m}^{\mathsf{T}}) \right] \quad \text{Posterior covariance.}
\end{aligned}
\tag{2.17}
$$

Figure 2.5 illustrates the summary of the SKF filtering for two regime switching state variable $s_t \in \{1, 2\}$ representing 1: stationary and 2: non-stationary regimes.

2.2.3 Parameter Estimation

In BDLM, each model dynamic involves a parameter set $\mathcal{P}$ whose values need to be estimated from a training dataset. The parameter set consists in the transition prior probabilities $Z^{ii}, \forall i = 1, 2$ and the parameters associated with model matrices $\{\mathbf{A}_t, \mathbf{C}_t, \mathbf{Q}_t, \mathbf{R}_t\}$. In the case of multiple observations, additional parameters are defined in order to establish dependencies between the observations. These parameters will be further discussed in Section 3. It is common to use the *maximum likelihood estimation* (MLE) technique in order to estimate the
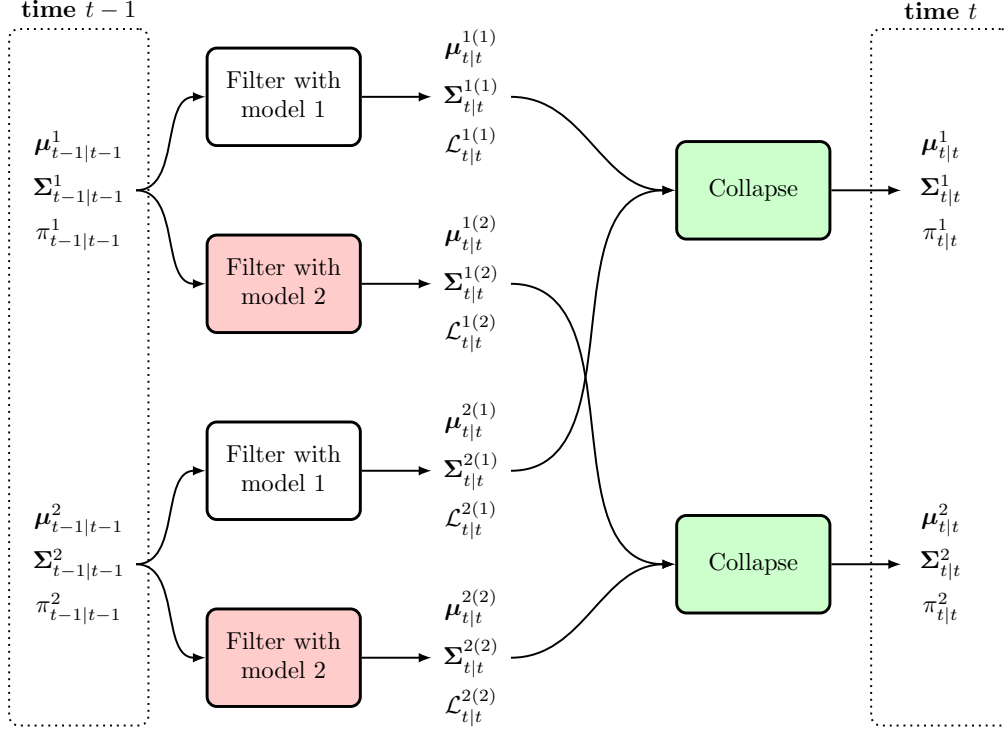
Figure 2.5 Illustration of the SKF algorithm for the two regime models 1: stationary and 2: non-stationary. The superscript $^{(\cdot)}$ indicates the corresponding regime's model that being used for the computation. The Figure is reproduced from Nguyen and Goulet [2].

parameters, where the objective is to obtain the *optimal parameter set*

$$\mathcal{P}^* = \arg\max_{\mathcal{P}} \ \ln f(\mathbf{y}_{1:\mathrm{T}}|\mathcal{P}), \tag{2.18}$$

which maximizes the joint log-likelihood $\ln f(\mathbf{y}_{1:\mathrm{T}}|\mathcal{P})$. The joint log-likelihood is obtained by the summation of the marginal log-likelihoods at each time t following

$$\ln f(\mathbf{y}_{1:\mathrm{T}}|\mathcal{P}) = \sum_t \ln f(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \mathcal{P}), \tag{2.19}$$

$$= \sum_t \ln \left[\sum_j \sum_i \ \mathcal{L}_t^{i(j)} \cdot Z^{i(j)} \cdot \pi_{t-1|t-1}^j \right],$$

where the marginal likelihood $\mathcal{L}_t^{i(j)}$ corresponds to the joint prior probability of observations of the model classes following

$$\begin{aligned}
\mathcal{L}_t^{i(j)} &= p(\mathbf{y}_t | S_t = j, S_{t-1} = i, \mathbf{y}_{1:t-1}, \mathcal{P}), &(2.20)\\
&= \mathcal{N}\left(\mathbf{y}_t; \mathbf{C}_t^j(\mathcal{P})\boldsymbol{\mu}_{t|t-1}, \mathbf{C}_t^j(\mathcal{P})\boldsymbol{\Sigma}_{t|t-1}^{i(j)}\mathbf{C}_t^j(\mathcal{P})^\intercal + \mathbf{R}_t^j(\mathcal{P})\right).
\end{aligned}$$

The maximization of the joint log-likelihood in Equation 2.18 is typically carried on using gradient-based optimization algorithms [55].

2.2.4 Example of Application

In order to illustrate the application of the SKF for detecting anomalies, consider Figure 2.6 where 6 years of observations are collected on the Champlain Bridge in Canada [56]. Figure 2.6a shows the elongation of an optical fiber sensor on the bridge that carries reversible and irreversible responses of the bridge over time. In addition, Figure 2.6b shows the air temperature of the area near the bridge.
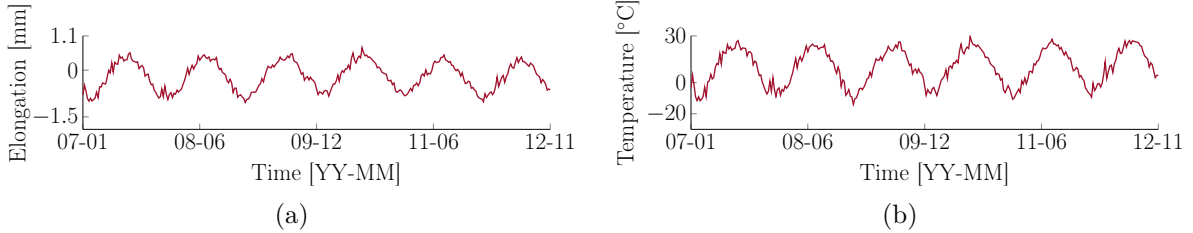
Figure 2.6 Observation data for a bridge in Canada: (a) Elongation in mm and (b) Temperature in °C.

Figure 2.7 illustrates the hidden state decomposition carried out by a BDLM using the regime switching technique presented in Section 2.2.2. Figures 2.7a-2.7c show the irreversible responses of the structure due to changes in degradation's kinematics. Figures 2.7d and 2.7e show the reversible responses of the structure. In particular, Figure 2.7d shows the external effects that are not caused by the temperature and Figure 2.7e illustrates the external effects due to observed temperature data. Note that Figure 2.7 does not show the uncertainty around the estimations. The stationary and non-stationary regimes represent the normal and abnormal conditions of the structure, respectively. Figure 2.7f shows the probability of the non-stationary regime for the same data presented in Figure 2.6a. The probability of the non-stationary regime is around zero for all times except the time, when the probability of the non-stationary regime suddenly increases to one. This indicates the presence of an anomaly
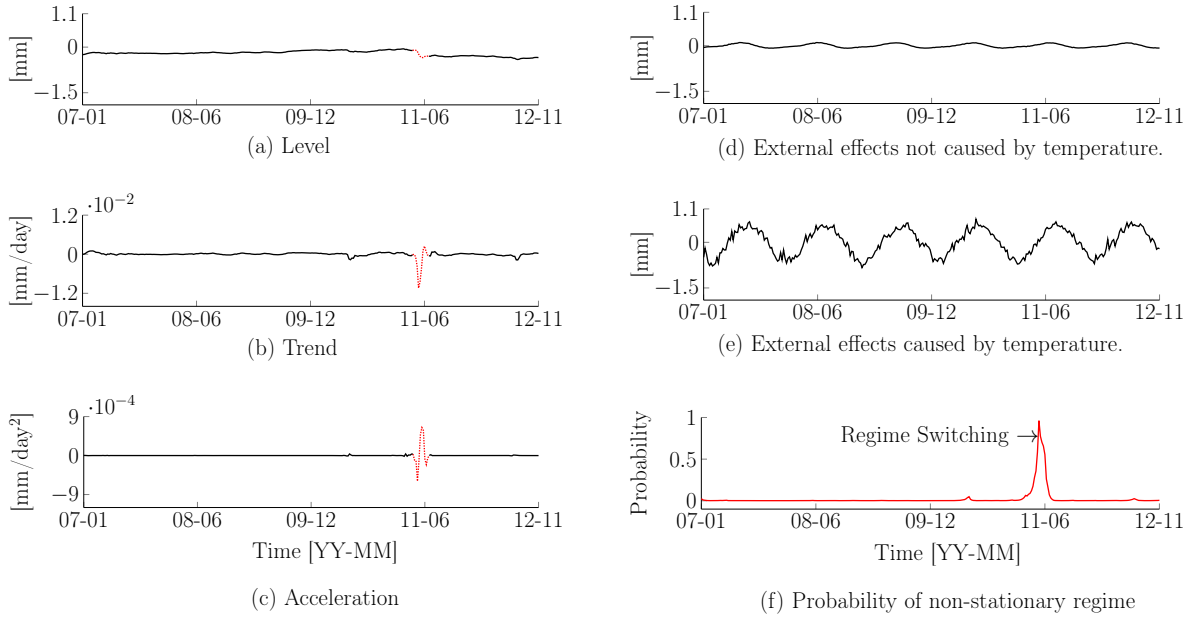
Figure 2.7 Illustration of decomposed observation data. Red dotted lines are corresponding to the detected anomaly period. The solid black lines indicate the mean value of the estimated hidden states. For a better illustration, the uncertainty around the mean values are not displayed.

during the corresponding timestamps as shown by red dotted lines in Figures 2.7a-2.7c. The probability of the regime switching, as shown in Figure 2.7f, can be used for decision making. For instance, when the probability of regime switching is close to one, a decision maker has a strong indication that a structure is in abnormal condition and that further engineering attention is needed.

2.2.5   Limitations

The main limitation of making decisions solely based on the probability of the non-stationary regime is that it is prone to false or missed alarms, especially for low probability events; It is common to have situations where the probability of the abnormal regime is low, while the underlying irreversible responses indicate a possible switch between two regimes, and vice versa. Therefore, the decision to trigger an alarm or not needs to consider not only the probability of regime switching, but also other criteria such as changes in the degradation's speed. To illustrate this point, consider Figure 2.8 where the irreversible responses and the probability of the non-stationary regime for a structural response is obtained. Two interesting

(a) Level

(c) Acceleration

(b) Trend

(d) Probability of non-stationary regime
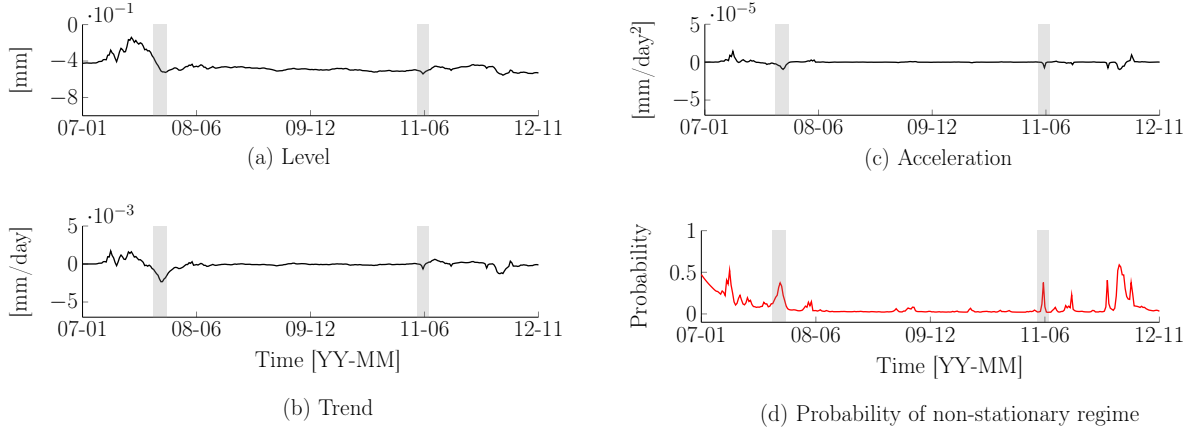
Figure 2.8 Illustration of the probability of the non-stationary regime and hidden states: shaded areas are corresponding to timestamps with same probability of anomaly. The solid black lines indicate the mean value of the estimated hidden states. For a better illustration, the uncertainty around the mean values are not displayed.

periods are emphasized by gray shaded areas, where the probability of the non-stationary regime is close to 40%. Decision making based on a predefined probability value, i.e., a threshold value, for these two periods perhaps results in the same action, i.e., triggering an alarm. However, the decision maker cannot be certain whether triggering an alarm is rightful or not. To overcome this limitation, there is a need for leveraging information from both degradation's speed and the probability of regime switching. In Figures 2.8a–c, although the two timestamps have identical probability of anomalies, the estimated hidden states are different. Such a difference can be used to improve the decision making. Later in Section 5.4, we will further study this limitation by quantifying probabilistically the anomaly detectability when relying only on a threshold value corresponding to the probabilities of a non-stationary regime.

Using only the degradation's speed and the probability of regime switching is not sufficient for decision making because the anomaly characteristics are not modeled in the analysis. The anomaly's magnitude and its duration affect the probability of regime switching and degradation's kinematics. However, there is no mechanism to identify these characteristics based on a BDLM. As a result, quantifying the anomaly detectability and subsequently associating the magnitude of detected anomalies to the severity levels of structural damages are not possible. The disciplines of reinforcement learning and imitation learning are suited for addressing their limitations.

2.3 Machine Learning

In general, anomaly detection methods encompass three areas of machine learning, namely *supervised learning*, *unsupervised learning*, and *reinforcement learning* [38]. In supervised learning, the model receives a sequence of inputs and the corresponding desired outputs such as labeled observation data indicating the normal and abnormal condition of structures. In such a setup, the goal is to correctly produce new outputs for new given inputs [19,57]. On the other hand, in unsupervised learning, the model receives a sequence of inputs without labeled outputs. The goal is then to build a model that is a representation of input patterns [30,32].

2.3.1 Reinforcement Learning

In reinforcement learning (RL), there is no sequence of inputs either to be mapped to desired outputs like in the case of supervised learning or to be examined to find patterns like in the case of unsupervised learning. Instead, it enables a decision maker, called the *agent*, to make decisions by taking *actions* in an *environment* in order to achieve its goal [58, 59]; RL is a sequential decision making process, where an agent interacts with its environment by taking actions that influence the environment's state. Accordingly, the environment gives a feedback in the form of a *reward*, a signal indicating whether the agent takes beneficial actions. The agent interacts with its environment in order to learn an *optimal policy* consisting in the best actions to take as a function of the state of the environment.

Applications of RL are typically seen in game playing such as the Atari environment [60] and the Go game [61]. In these cases, the environment is an emulator allowing an agent to take some actions defined within the game. For example, take the Breakout game shown in Figure 2.9; The game consists in some brick layers, a ball, and a paddle. The agent uses the paddle to hit the ball into the bricks in order to eliminate them. The agent can move the paddle to the left or right in order to control the ball and eventually remove all the bricks.
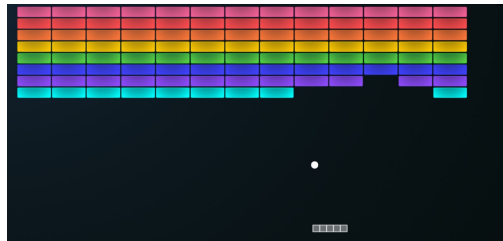
Figure 2.9 Illustration of the Breakout Atari game emulator. It consists in some layers of bricks, a paddle, and a ball. The agent moves the paddle to the left or right in order to hit the ball into the bricks and eliminate them. The figure is adapted from Yandex LLC. [3].

If the ball hits the floor, the agent would lose the game and the emulator would restart the game from the beginning: At each time step, the agent selects an action from a set of actions $\{\text{left}, \text{right}\}$ in order to move the ball. The action selected is passed to the emulator which modifies the state of the ball and returns its new location. In addition, the emulator gives a reward in the form of a scoring system based on the number of bricks eliminated. The agent's objective is to maximize the long-term accumulation of these scores during the game without losing the ball.

In the beginning, the agent might lose the ball often. Eventually, as the agent learns how to play the game, it reaches the maximum score. The learning process in the context of the RL involves two important aspects; The first aspect is that the agent uses different sources of information from the emulator in order to form the environment's states to interact with. For instance, in the original work done by Mnih et al. [60], the sources of information for the Breakout game are the Atari frames consisting in 210×160 pixels color images. The second aspect of the learning is associated with the consideration of the long-term effects of actions. For instance, in the Breakout game, the agent requires not only to collect the current reward by choosing the right move, but also it needs to take into account the next moves that enable the agent to collect scores for new ball positions; For example, the agent might need to bounce the ball by moving to the right instead of left. Because, in some time steps after, the ball position is on the right side of the screen, and therefore, the agent can bounce the ball back again without losing it.

In the context of SHM, the game should be replaced by the anomaly detection procedure, and actions by decisions based on triggering an alarm or not. In this setup, RL offers solutions to the limitations of the SKF-BDLM, presented in Section 2.2.2, by considering the temporal aspects related to the degradation's speed and the time required by an agent to detect anomalies: RL-based methods enable the decision making process to account for different sources of information in order to build the environment that the agent interacts with. For instance, the agent can incorporate information from the SKF-BDLM analysis including the regime switching probabilities and the degradation's speed. In addition, the formalism of RL accounts for the long-term accumulation of rewards in order to potentially delay or speed up a decision. In the context of anomaly detection, the long-term planning considerations of RL controls the time to detection of anomalies. Therefore, the RL can enable an agent to account for early anomaly detection as well as preventing false alarms. This section reviews the elements of RL and explores existing RL algorithms available in the literature.

Elements of Reinforcement Learning

RL problems involve seven primary elements [62], namely, *agent, environment, state, policy and action, reward, (action-)value function*, and, optionally, a *model*. These elements are formalized via a Markov decision process (MDP) [49]: for a discrete time step set $t = \{1, 2, 3, \ldots\}$, a MDP is defined by a tuple $\langle \mathcal{S}, \mathcal{A}, r_t, p(\mathbf{s}_{t+1} | \mathbf{s}_t, a_t), \gamma \rangle$. Here, $\mathcal{S} \subseteq \mathbb{R}^{\mathtt{S}}$ is a $\mathtt{S}$-dimensional continuous environment, where $\mathtt{S} \in \mathbb{N}$ is the number of dimensions; The action-space $\mathcal{A}$ is a discrete set containing $\mathtt{A} \in \mathbb{N}$ actions; At time t, the immediate reward $r_t \equiv r(\mathbf{s}_t, a_t) \in \mathbb{R}$ is a function of the corresponding state $\mathbf{s}_t \in \mathcal{S}$ and action $a_t \in \mathcal{A}$; $p(\mathbf{s}_{t+1} | \mathbf{s}_t, a_t) \in (0, 1)$ is the probability to transit from the state $\mathbf{s}_t$ to the successor state $\mathbf{s}_{t+1}$ after taking the action a_t; and $\gamma \in (0, 1]$ is the discount factor. A policy $\pi(\mathbf{s}_t) = a_t$ is a function defining the action that an agent takes given the state it is in. In the special case, where policies are *stationary* and deterministic, an agent that follows the policy $\pi(\mathbf{s}_t)$ always takes the same action for the same state.

In order to introduce RL concepts, consider the example of a robot moving on the ground where the goal is to reach a target and avoid obstacles. An *agent*, is an entity that observes the environment using sensors and takes actions in order to achieve its goal. In this example, the agent is the robot moving in different directions. The environment $\mathcal{S}$, is a state-space that the agent interacts with. For the moving robot example, the environment can consist in the surrounding areas and obstacles perceived using sensors. An *environment's state*, or simply a state, $\mathbf{s}_t \in \mathcal{S}$ at time t is a particular configuration of an environment that an agent can take. For example, in the case of a moving robot, the state can be its position in a two dimensional space. A policy $\pi(\mathbf{s}_t) = a_t$ is a function of the state which represents the action $a_t \in \mathcal{A}$ that the agent takes for a given state $\mathbf{s}_t$. For a moving robot, the action $a_t \in \mathcal{A} = \{\text{left}, \text{right}, \text{forward}, \text{backward}\}$ can be selected from a set consisting in movements along different directions. A reward $r(\mathbf{s}_t, a_t)$ is a scalar that an agent receives for being in a state $\mathbf{s}_t$ and taking an action a_t; it defines what is a good and bad action for the agent, given the state it was in. For a moving robot, the reward can be a positive scalar when the agent reaches the target and a negative one when it hits the obstacles. An action-value function $q_\pi(\mathbf{s}_t, a_t)$, for an agent being in the state $\mathbf{s}_t$ at time t following the policy π, is defined in terms of the expected value of the *utility* $\mathbb{U}(\mathbf{s}_t, \pi) = r_t(\mathbf{s}_t, a_t) + \sum_{t'=t+1}^{\infty} \gamma^{t'-1} r_{t'}(\boldsymbol{S}_{t'}, a_{t'})$, describing the total accumulated discounted rewards that an agent receives when following the policy.

It is common to define the action-value function as

$$q_\pi(\mathbf{s}_t, a_t) = \mathbb{E}_\pi \left[\mathbb{U}(\mathbf{s}_t, \pi) \right] \tag{2.21}$$

$$= \underbrace{r(\mathbf{s}_t, a_t)}_{\text{immediate reward}} + \gamma \cdot \overbrace{q_\pi(\boldsymbol{S}_{t'}, a_{t'})}^{\text{successor action-value function}}.$$

In Equation 2.21, for an agent being in the state $\mathbf{s}_t$ at the time t, the successor state $\boldsymbol{S}_{t'}$ is a random variable where $t' = t+1$. In addition, $\gamma \in (0, 1]$ is the *discount factor* quantifying how much future reward values are discounted in comparison with the current one. In this view, an action-value is evaluated by not only collecting the reward associated with the current action, but also considering future rewards due to succeeding actions: The action-value of a state-action pair $(\mathbf{s}_t, a_t)$ at time t is the expected return of the agent that starts from the state $\mathbf{s}_t$ and takes the action a_t, and takes the future actions by following the policy π [62]. Also, the action-value function can be expressed based on the immediate reward and the action-value function of the successor state as shown in the second line of Equation 2.21. The optimal stationary policy $\pi^*(\mathbf{s})$ is then obtained by the maximization of the action-value function following

$$\pi^*(\mathbf{s}) = \arg\max_{a_t} q(\mathbf{s}_t, a_t). \tag{2.22}$$

In the context of RL, a model is defined in the form of *transition probabilities* $p(\boldsymbol{S}_{t+1}|\mathbf{s}_t, a_t)$ indicating the probability that the agent transits from one state to another by taking an action. Figure 2.10 shows schematically the interaction between an agent and an environment. Being in the state $\mathbf{s} \equiv \mathbf{s}_t$, the agent take the action $a \equiv a_t$. The environment returns the scalar reward $r(\mathbf{s}_t, a_t)$ and the agent transmits to the state $\boldsymbol{S}' \equiv \boldsymbol{S}_{t+1}$ with the probability $p(\boldsymbol{S}'|\mathbf{s}, a)$. This procedure continues until the policy converges or there is a terminal state such as reaching the target for the moving robot example.
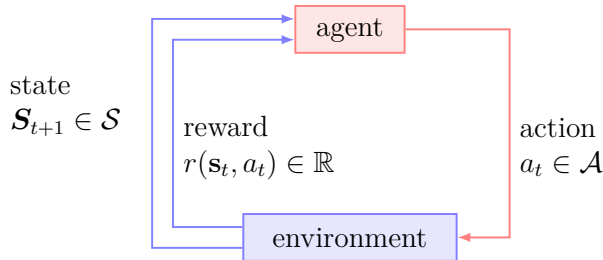
Figure 2.10 The agent-environment interaction in reinforcement learning.

RL Algorithms

Solving a MDP requires transition probabilities, which are not always available. In this regard, RL algorithms are divided into two categories, namely *model-based* or *model-free*. In model-based RL, the transition probability $p(\boldsymbol{S'}|\mathbf{s}, a)$ for the agent to transit from the state $\mathbf{s}$ to $\boldsymbol{S'}$ due to the action a is known. In model-based RL, Equation 2.21 leads to the Bellman equation [63] following

$$q_\pi(\mathbf{s}, a) = r(\mathbf{s}, a) + \gamma \sum_{\boldsymbol{S'}} \underbrace{p(\boldsymbol{S'}|\mathbf{s}, a)}_{\text{RL model}} q_\pi(\boldsymbol{S'}, a'). \tag{2.23}$$

The Bellman equation can be solved using the value or policy iteration algorithms [58]. Alternatively, *temporal difference* methods, known as TD(λ), have been introduced by Sutton [59] in order to cope with situations where transition probabilities are not known. In these approaches, the value function is evaluated between states $\mathbf{s}_t$ and $\mathbf{s}_\lambda$ representing the current state and the successor λ-state after the current one, respectively. The difference between two values is then used to update the action-value corresponding to the state $\mathbf{s}_t$. The simplest form of temporal difference method TD($\lambda = 0$) is *Q-learning*, where the action-value functions known as *Q-value functions*, are evaluated between two immediate successor states $\mathbf{s}$ and $\boldsymbol{S'}$.

When an environment is discrete, conventional RL algorithms such as the value and policy iteration or Q-learning are used. When an environment is continuous, one can discretize the environment and use conventional RL algorithms, or employ *function approximation* methods. In general, function approximation methods are developed to either estimate action-values and find the optimal policy according to Equation 2.22 or directly approximate the optimal policy [62]. A function approximator receives the state $\mathbf{s}_t$ at time t as an input and returns approximated action-value set $\{\hat{q}(\mathbf{s}_t, a_{n,t}; \boldsymbol{\theta}), \forall n = 1 : \mathtt{A}\}$, where, $a_{n,t} \in \mathcal{A}$ is the nth element of the action set $\mathcal{A}$ containing $\mathtt{A}$ individual actions. In addition, the function approximator is characterized by a set of parameters $\boldsymbol{\theta}$ that is learned from agent-environment interactions. Once action-values are estimated, the optimal policy can be determined from Equation 2.22.

Different types of linear and nonlinear function approximation methods are listed in [62, 64, 65]. Recent advancements in deep neural networks have led to the development of efficient and reliable approximation methods such as deep Q-learning and actor-critic methods [61]. The primary limitation of function approximation methods is that the convergence is only guaranteed in case of linear function approximation, where the convergence in nonlinear cases is an active field of research in the machine learning community [66].

Q-Learning

The main obstacle in solving Equation 2.23 is to have access to the transition probabilities. If such transition probabilities are not known, we cannot directly maximize the action-value functions in order to obtain optimal policies. Instead, an agent estimates the action-values while interacting with its environment [67]. The action-value function is referred as *Q-value function* $Q(\mathbf{s}, a)$, and is evaluated through the updating equation

$$Q(\mathbf{s}_t, a_t) := Q(\mathbf{s}_t, a_t) + \alpha_t \left[r(\mathbf{s}_t, a_t) + \gamma \max_{a_{t+1}} Q(\boldsymbol{S}_{t+1}, a_{t+1}) - Q(\mathbf{s}_t, a_t) \right], \qquad (2.24)$$

where $:=$ is the *update symbol*; $Q(\mathbf{s}_t, a_t)$ is the updated Q-value at time t; $\alpha_t \in (0, 1)$ is the learning rate; and $\mathbf{s}_t$ is a vector-valued state of the agent. Equation 2.24 converges to the optimal value function asymptotically [68, 69], where it updates one state-action pair $(\mathbf{s}_t, a_t)$ at each time-step t according to the current reward $r(\mathbf{s}_t, a_t)$, discount factor γ, learning rate α_t, and the expected maximum Q-value of the successor state $\boldsymbol{S}_{t+1}$.

Intuitively, an agent should learn less after having spent more time experiencing an environment. Therefore, one can start with a high learning rate value to allow fast changes and reduce it as time progresses. It is common to use an iteration-dependent *learning rate* function such as

$$\alpha_t = \alpha(\mathbb{N}_{1:t}(\mathbf{s}_t, a_t); c_\alpha) = \frac{c_\alpha}{c_\alpha + \mathbb{N}_{1:t}(\mathbf{s}_t, a_t)}, \qquad (2.25)$$

where $\mathbb{N}_{1:t}(\mathbf{s}_t, a_t)$ is the number of times a state-action pair has been visited up to time t and $c_\alpha \in \mathbb{R}^+$ is the *learning rate constant*. In addition, to ensure that the learning-rate can take any value and that Equation 2.24 converges, the learning-rate must satisfy $\sum_{t=1}^{\infty} \alpha_t = \infty$ and $\sum_{t=1}^{\infty} \alpha_t^2 < \infty$. The reader is referred to [70] for a description of different learning-rate functions and their influences on the learning procedure.

In order to identify which action should be selected at a time t, it is necessary to present the *exploitation* and *exploration* [71] dilemma; Exploitation uses the current knowledge about the state $\mathbf{s}_t$ to select the action a_t that maximizes the Q-value. In contrast, exploration selects other actions rather than the one that is currently thought to be the best in order to potentially reaches a solution associated with higher Q-values. Among different strategies, the ϵ-greedy algorithm [67] is a common approach to address the exploitation and exploration dilemma: At a time t, an agent explores by selecting an action randomly with the probability $\epsilon_t \in (0, 1)$ and exploits by selecting an action from the current policy with the probability $1 - \epsilon_t$. Intuitively, it is better to encourage an agent to explore more during the early stage of learning and then reduce the exploration probability as time evolves. It is common to define

an iteration-dependent exploration function as

$$\epsilon_t \equiv \epsilon(\mathtt{N}_{1:t}(\mathbf{s}_t); c_\epsilon) = \frac{c_\epsilon}{c_\epsilon + \mathtt{N}_{1:t}(\mathbf{s}_t)}, \tag{2.26}$$

where $\mathtt{N}_{1:t}(\mathbf{s}_t)$ is the number of times a state has been visited up to time t regardless of the selected actions, and $c_\epsilon \in \mathbb{R}^+$ is the *exploration constant*; To select the action a_t at time t, a sample u_t from a uniform distribution between zero and one, $\mathcal{U}(u_t; 0, 1)$, is drawn and compared with the corresponding exploration probability ϵ_t. In this setup, the action selection is expressed mathematically as

$$a_t = \begin{cases} \text{Random selection from } \mathcal{A}, & u_t \sim \mathcal{U}(u_t; 0, 1) \leq \epsilon_t \\ \underset{a_t}{\arg\max}\ Q(\mathbf{s}_t, a_t), & \text{Otherwise} \end{cases}. \tag{2.27}$$

The reader is referred to [62] and [72] for a review of different exploration strategies.

In the context of civil structures, reinforcement learning has been employed in the field of active structural control where a controller takes some actions based on a feedback from an instrumented structure. The goal is then to minimize an error signal defined as the difference between the actual and desired feedback from a system [73, 74]. Although some of definitions and principles in RL and control theory are different, the core idea behind both approaches remains the same. However, few researches are dedicated to precisely formulate structural control problems in the form of a RL framework [75, 76]. For instance, Khalatbarisoltani et al. [76] have employed Q-learning for an active control of a mass driven system on a laboratory-scale structure. For the environment, they have used a physics-based model based on a mass-spring-damper system, which introduces limitations as discussed in Section 2.1 with respect to using physics-based models. RL-based anomaly detection approaches have been seldom applied in SHM contexts despite being well suited for the problem at hand. For examples, Huang et al. [77] and Yu an Sun [24] have proposed a value-based and policy-based *forward* anomaly detection methods based on deep reinforcement learning. Oh and Iyengar [78] have used an *inverse* RL method for the task of anomaly detection, where the objective is to determine an underlying reward function, and subsequently the agent's preference, from a sample of optimal actions. The above-mentioned researchers have showed advantages of both forward and inverse RL for anomaly detection on various time series. However, the time series employed do not account for the specificities of SHM data that are collected on civil infrastructures. In addition, they have used performance metrics such as F1-score, accuracy, and recall, which are used to evaluate the performance of the model built in the context of classification where the objective is to discriminate between the normal and

abnormal states of structures. However, they do not account for anomaly characteristics, i.e., the anomaly magnitude, its duration, and the time required to detect an anomaly, that are continuous variables.

2.3.2 Imitation Learning

One of the limitations in the context of reinforcement learning is the requirement to design a reward function $r(\mathbf{s}_t, a_t)$ that encourages or prevents an agent to take actions in order to accomplish the task in hand. To overcome this limitation, Pomerleau [79] proposed an approach that consists in learning from demonstrations, also known as imitation learning (IL). In IL, an expert, usually a human, provides the optimal policy within a set of demonstrations associated with the desired behavior. Then, the agent learns the optimal policy by imitating the expert's policy. For example, in the game Breakout presented in Section 2.3.1, an expert assigns optimal actions for the set {left, right} for moving the paddle in different situations, instead of designing a reward function in order to encourage an agent to move left or right.

There are two approaches towards imitation learning in the literature. In the first approach, an agent learns to directly imitate the expert's policy. The simplest form of this approach is called *behavior cloning* which is a supervised learning approach. Note that in this thesis we use the term *imitation learning* instead of *supervised learning* in order to follow the existing terminology used in the literature. In this setup, a classifier is trained from a demonstrations set in order to predict the expert's behavior. The training set is obtained from an expert's policy and it consists in state-action pairs. The behavior cloning approach is shown to yield poor results when the demonstrations issued by the expert are not uniformly sampled from the environment. Ross et al. [80] introduced an algorithm known as *data aggregation* (DAgger) in order to relax this limitation. As such, demonstrations are used incrementally whenever needed; An agent learns to imitate the expert's policy with given demonstrations, and whenever there are states which are not in the current demonstrations set, the agent makes a query about optimal actions from the expert for these new states. The reader is referred to Hussein et al. [81] for a comprehensive review of existing methods relying on directly imitating the expert's policy. In the second approach, which is not considered in the current work, an agent learns to indirectly imitate the expert's policy by first learning the reward function and then learning the optimal policy. This approach is known as inverse reinforcement learning and the reader is referred to [82] for detailed review of existing methods.

The elements of the imitation learning approach are similar to the ones of the reinforcement learning approach presented in Section 4.1 with respect to an environment $\mathcal{S}$, an action

space $\mathcal{A}$, a transition probability between two consecutive states given the action taken by an agent $p(\boldsymbol{S}_{t+1}|\mathbf{s}_t, a_t)$, and a policy $\pi(\mathbf{s}) = a_t$, a function that determines the action taken by an agent for a given state $\mathbf{s}_t$. In addition, the expert's policy $\pi^{\text{expert}}(\mathbf{s}) = a_t^{\text{expert}}$ is a function defined by an expert, that determines the optimal action for a given state $\mathbf{s}_t$. Moreover, a demonstrations set Ξ is defined as

$$\Xi = \{\hat{\mathcal{Z}}_1, \hat{\mathcal{Z}}_2, \cdots, \hat{\mathcal{Z}}_{\text{Z}}\}, \tag{2.28}$$

where, it contains Z demonstrations $\hat{\mathcal{Z}}$, each consisting in state-action pairs which issued by an expert, i.e., $\hat{\mathcal{Z}}_i = \{(\mathbf{s}_1, a_1^{\text{expert}}), (\mathbf{s}_2, a_2^{\text{expert}}), \cdots\}$. In this setup, the imitation learning consists in an optimization problem following

$$\pi^*(\mathbf{s}) = \arg\min_{a_t} \sum_{\mathbf{s} \in \Xi} J(\pi(\mathbf{s}), \pi^{\text{expert}}(\mathbf{s}); \boldsymbol{\theta}), \tag{2.29}$$

where, $\pi^*(\mathbf{s})$ is the optimal policy, and $J(.)$ is a loss function parameterized by $\boldsymbol{\theta}$, a set of unknown parameters associated with the classification model [55]. Solving Equation 2.29 leads to a classifier predicting the optimal actions that an agent should take given the states.

2.4 Conclusion

Most existing SHM methods rely on threshold-based approaches for detecting anomalies. These approaches are limited by their poor control over false alarms which jeopardize the viability of the anomaly detection system. In most cases, existing anomaly detection methods focus on detecting specific anomalies instead of considering a wide range of possible ones. Moreover, many existing SHM methods for detecting anomalies are applied on laboratory structures, yet their applications on real-world structures subject to operational conditions have not been examined.

The methods based on Bayesian dynamic linear models with regime switching, i.e., SKF-BDLM, offer a part of the solution to the limitations associated with threshold-based anomaly detection methods. Using a regime switching paradigm enables having control over false alarms while maintaining a high anomaly detectability. In addition, BDLM methods have been applied on full-scale structures that are operating under real-life conditions [52, 56, 83]. Therefore, these methods are suitable for long-term monitoring of structures. However, the main issue with SKF-BDLMs is that there is no systematic criteria for triggering alarms that would consider the temporal aspects related to the degradation's speed and the time to detection; The only source of information for anomaly detection is the probability of regime

switching, which in many cases, is not sufficient for a robust anomaly detection. In addition, BDLM methods lack a mechanism in order to take into account delayed and early detections. For example, delaying a decision regarding the detection of anomalies can help preventing false alarms.

In order to identify an efficient sensor configuration that is able to discriminate various structural damages, we need to develop a systematic procedure along with metrics such as annual false positive ratio and the probability of true positive detections. Such a systematic procedure should be probabilistic and requires characterizing the diversity of possible anomalies through their magnitude, duration and time to detection. Developing such a metric requires an anomaly detection method that incorporates the degradation's speed and characterizes delayed and early detections.

CHAPTER 3   Jacques-Cartier Bridge Monitoring and Data Analaysis

This chapter first presents the Jacques Cartier Bridge and its general structural characteristics. In addition, it introduces the monitoring campaign along with the instrumentation configuration including sensor types and locations. Furthermore, this chapter illustrates the data acquired and discusses the issues identified for some of the sensors. Finally, this chapter provides the empirical models built for each sensor using the Bayesian dynamic linear models described in Section 2.2.

3.1   Bridge Presentation

The Jacques Cartier Bridge is a steel cantilever truss bridge crossing the Saint-Lawrence river, Canada. Originally inaugurated in May 1930, this 2765 m bridge connects Montreal to Longueuil. Figure 3.1 illustrates the schematic elevation of the Jacques Cartier Bridge indicating nine distinct sections differing in the types of sub- and superstructures. The structure of the bridge consists in: (i) concrete for the deck and substructure of the south shore approach, and (ii) steel for the superstructure and substructure of the Montreal approach. Each of the 40 spans is composed of a 23.1 m wide concrete deck supported by riveted trusses, and dowels, as well as tension anchors in the spans associated with the sections 3 and 7.
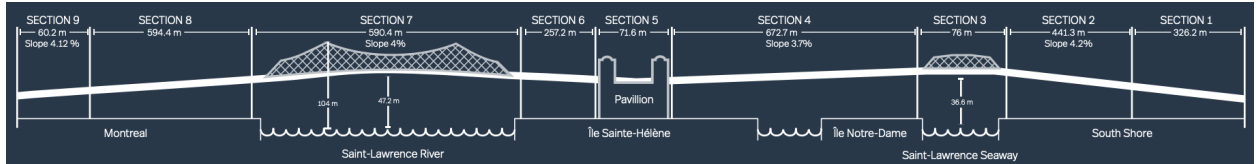
Figure 3.1 Illustration of the not-to-scale schematic elevation (upstream view) of the Jacques Cartier Bridge. The bridge has 40 spans grouped in 9 different sections. The figure is reproduced from the Jacques Cartier Bridge technical data sheet [4].

The Société des ponts Jacques Cartiers et Champlain has been responsible for managing various interventions to ensure the bridge's safety and for extending its life span. Interventions encompass: (i) major renovation programs such as replacing the whole bridge's deck in 2001−2002, and (ii) a structural steel reinforcement program involving reinforcing and replacing steel ribs and plates. In the next sections, we present the monitoring campaign and measurement system layout that has been installed on the Jacques Cartier Bridge for the purpose of supporting the methodological developments in this thesis.

3.2 Monitoring Campaign

Starting in 2019, three spans have been instrumented on the Longueuil approach, see the section 2 region in Figure 3.1. These spans are located between the axes 4 and 7 as depicted by the shaded green area in Figure 3.2. We refer to these spans as "4 − 5", "5 − 6", and "6 − 7". Except for repairs due to the structural steel reinforcement program, the three spans are identical with respect to their superstructures, geometry, and boundary conditions; Each span is supported by fixed supports on the north side, and expansion joints and bearing devices on the south side. In Figure 3.2, the fixed supports and expansion joints are indicated by F and EXP, respectively.
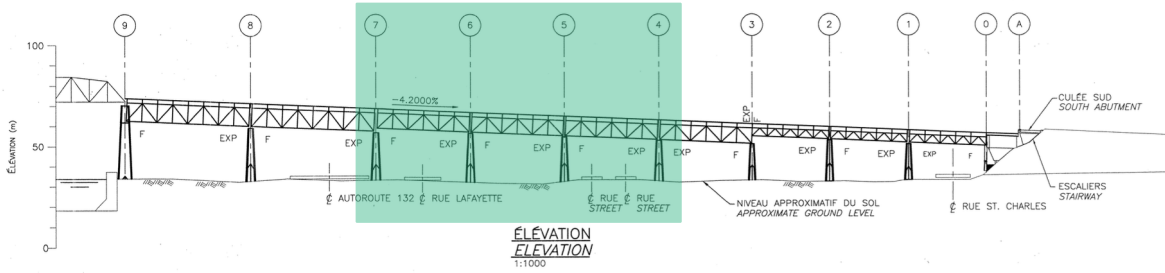
Figure 3.2 Illustration of the selected spans (shaded are) between the axes 4 to 7. The drawing is reproduced from the Jacques Cartier Bridge deck replacement program documentation [5].

3.3 Instrumentation

The selected spans of the Jacques Cartier Bridge are instrumented with three types of sensors collecting measurements, that is inclination in degree [d°] for inclinometers, elongation in [mm] for extensometers, and air temperature in degree Celsius [C°] for thermometers. The instrumentation configuration for each span is identical and consists in three inclinometers, an extensometer, and a thermometer, which are installed on the downstream side of the span. Therefore, in total, there are nine inclinometers, three extensometers, and three thermometers. Figure 3.3 shows the location of the sensors on a typical span of the bridge. The drawing shows seven vertical axes named from 1 to 7. The two longitudinal inclinometers are located on the axes 1 and 3 and measure rotations around the y axis; the lateral inclinometer is located on the axis 3 and measure rotations around the x axis; the extensometer is located on the axis 4 on the lower cord measuring elongations along the x axis; the thermometer is located on the axis 4. Note that accelerometers have been installed, however, the data are not considered because the signal-to-noise ratio is too low to make a distinction between
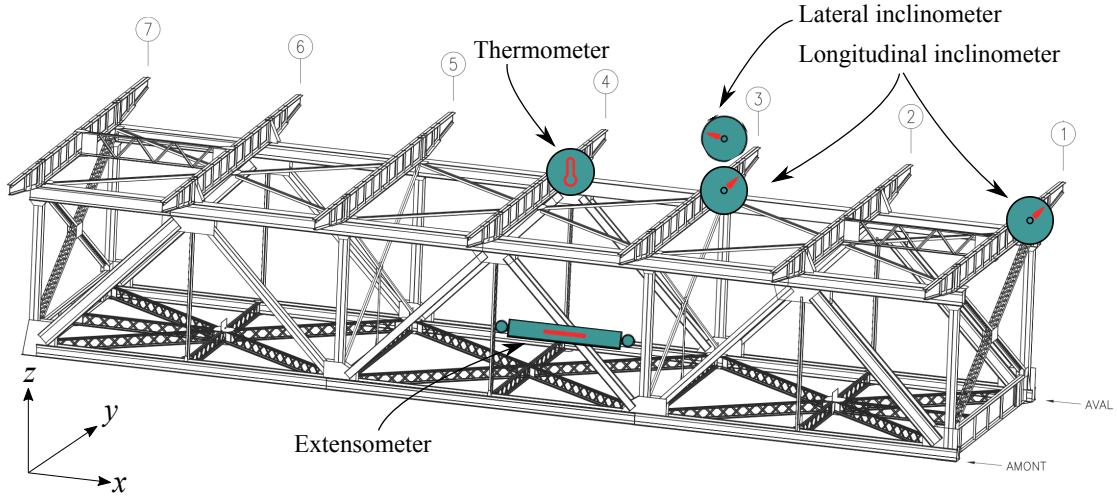
Figure 3.3 Sensor configuration for each span: an extensometer on the axis 4, two longitudinal inclinometers on the axes 1 and 3, and a lateral inclinometer on the axis 3. All the sensors are installed on the downstream side of the span. The drawing is reproduced from the Jacques Cartier bridge steel repairs program documentation [6].

actual vibrations of the structure and the ones induced by external excitements such as traffic loading and wind. Each sensor has a particular ID indicating the sensor type and location as shown in Table 3.1. The third column in Table 3.1 shows the sensors' complete ID received from the consultant's data acquisition servers. For brevity, we define a generic simplified ID for each sensor. This allows to readily identify a sensor's type and location. Each simplified name consists in the span name followed by an abbreviation, that is I_1 for a longitudinal inclinometer on the axis 1, I_3 for a longitudinal inclinometer on the axis 3, I_L for a lateral inclinometer, E for an extensometer, and T for a thermometer. For instance, the complete name of the longitudinal inclinometer on the axis 1 on the span $4 - 5$, shown in the first row of the table, is "INC_LONG_4_5AV_1", which is simplified to "$4 - 5 : I_1$".

Table 3.1 Sensors' complete and simplified IDs.

| Sensor type | Span | ID | |
		Complete	Simplified
Longitudinal Inclinometer (axis 1)	$4-5$	`INC_LONG_4_5AV_1`	$4-5:\mathtt{I}_1$
	$5-6$	`INC_LONG_5_6AV_1`	$5-6:\mathtt{I}_1$
	$6-7$	`INC_LONG_6_7AV_1`	$6-7:\mathtt{I}_1$
Longitudinal Inclinometer (axis 3)	$4-5$	`INC_LONG_4_5AV_3`	$4-5:\mathtt{I}_3$
	$5-6$	`INC_LONG_5_6AV_3`	$5-6:\mathtt{I}_3$
	$6-7$	`INC_LONG_6_7AV_3`	$6-7:\mathtt{I}_3$
Lateral Inclinometer (axis 3)	$4-5$	`INC_LATE_4_5AV_3`	$4-5:\mathtt{I}_{\mathrm{L}}$
	$5-6$	`INC_LATE_5_6AV_3`	$5-6:\mathtt{I}_{\mathrm{L}}$
	$6-7$	`INC_LATE_6_7AV_3`	$6-7:\mathtt{I}_{\mathrm{L}}$
Extensometer (axis 4)	$4-5$	`EXT_4_5AV_4`	$4-5:\mathtt{E}$
	$5-6$	`EXT_5_6AV_4`	$5-6:\mathtt{E}$
	$6-7$	`EXT_6_7AV_4`	$6-7:\mathtt{E}$
thermometer (axis 4)	$4-5$	`TH_04_5AV_3`	$4-5:\mathtt{T}$
	$5-6$	`TH_05_6AV_3`	$5-6:\mathtt{T}$
	$6-7$	`TH_06_7AV_3`	$6-7:\mathtt{T}$

3.4 Data

The data acquisition started in May 2019 with the averaged acquisition frequency of one data point per 10 min. This research analyzes the data from August 2019 until November 2021; We have discarded the first three months of the data due to multiple sensor calibration issues that resulted in missing or improperly collected measurements. Figures 3.4a and 3.4b illustrate an example of the extensometer and temperature measurements on the span $5-6$ with the respective labels $5-6:\mathtt{E}$ and $5-6:\mathtt{T}$, according to Table 3.1. We use weekly
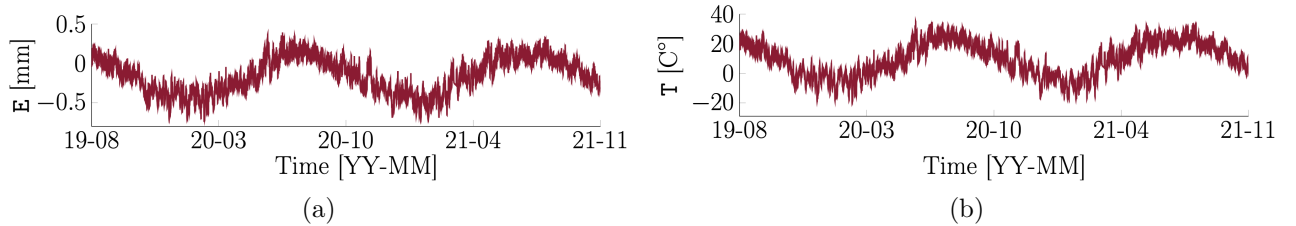
Figure 3.4 Illustration of the elongation and temperature data collected on the span $5-6$: (a) elongation data with an averaged period of acquisition of 10 min, and (b) raw temperature data with an averaged period of acquisition of 10 min.

averaged data for all measurements in the current research as it facilitates and speeds up the analysis without limiting the interpretation capacity. Figures 3.5a and 3.5b show the weekly averaged data corresponding to the raw data displayed in Figure 3.4a and 3.4b, where each data point on Figures 3.5a and 3.5b corresponds to the average of data points collected within a week.

(a)   (b)
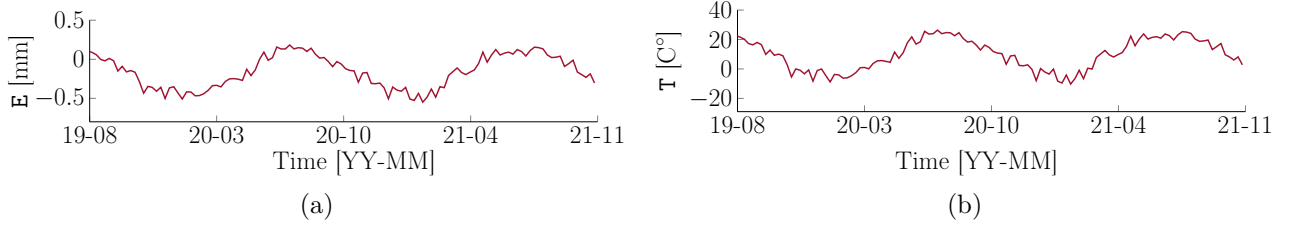
Figure 3.5 Illustration of the weekly averaged elongation and temperature data collected on the span $5 - 6$: (a) weekly averaged elongation data, and (b) weekly averaged temperature data.

The bridge responses measured by extensometers and inclinometers are affected by the temperature. Our study of these time series has revealed that the temperature is the dominant external effect. Figure 3.6 shows two examples of dependencies between the normalized weekly averaged temperature and elongation data as well as the normalized longitudinal inclination data. The normalization is carried out ensuring the data range is $(0, 1)$. Note that the normalization is linear and plotting the data in the original space would yield the same results. Here, such normalization is carried out for a better illustration resulting in

(a)   (b)
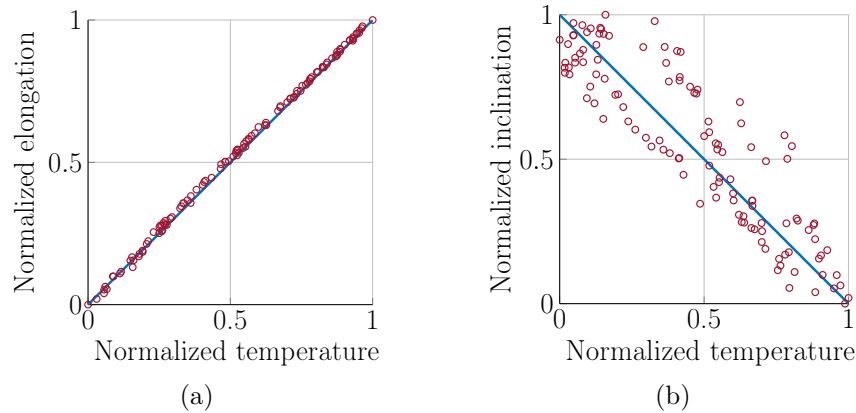
Figure 3.6 Illustration of the dependency between the normalized temperature data and the normalized structural measurements: (a) normalized elongation and (b) normalized inclination. The dependency between the temperature and elongation is positive, while it is negative for inclination.

figures with equal axes scales. Figure 3.6 displays a strong linear dependency between the temperature and the elongation and inclination data; The dependency between the temperature and elongation is positive, while it is negative for inclination. Figure 3.7 shows the weekly averaged data for each span. The gray areas represent removed timestamps for which the data was improperly collected or was missing. Moreover, the data corresponding to the sensor $6 - 7 : \mathtt{I}_3$ has not been collected after February 2021. The next section presents the analysis of these time series including the method used to account for dependencies between the temperature data and other measurements.

3.5   Results

This section presents the empirical models built for the three instrumented spans of the Jacques Cartier Bridge. We use Bayesian dynamic linear models (BDLM) in order to model the time series as described in Section 2.2. The modeling process encompasses decomposing the measurements by employing BDLM generic components, and creating dependencies between the structural responses and temperature data.

3.5.1   Jacques Cartier Bridge Time Series Modeling

For non-stationary regimes, we employ a level hidden state variable $x^{\mathtt{L}}$ in order to describe the *baseline* behavior of the structural response. The rate of change in the baseline is the trend hidden state variable $x^{\mathtt{T}}$. The rate of change in the trend is the local *acceleration* hidden state variable $x^{\mathtt{LA}}$. These three variables represent the degradation's kinematics describing the irreversible responses of the structure's degradation, the degradation's speed, and the degradation's acceleration. They jointly form the local acceleration vector $\mathbf{x}_t^{\mathtt{LA}} = [x_t^{\mathtt{L}}\ x_t^{\mathtt{T}}\ x_t^{\mathtt{LA}}]^{\mathsf{T}}$ similar to the one presented in Equation 2.4. For stationary regimes, the acceleration is zero for all the time steps. Therefore, the irreversible responses consist in a level and a local trend hidden state variables. They jointly form the local trend vector $\mathbf{x}_t^{\mathtt{LT}} = [x_t^{\mathtt{L}}\ x_t^{\mathtt{LT}}]^{\mathsf{T}}$ as presented in Equation 2.6. Note that the only irreversible response to model temperature data is the local level $x^{\mathtt{LL}}$ hidden state variable. The kernel regression vector $\mathbf{x}_t^{\mathtt{KR}}$ models the reversible *periodic* patterns for each observation [52]. For the measurements associated with inclinometers and thermometers, we use 10 hidden control points for the kernel regression which result in the vector of hidden states $\mathbf{x}_t^{\mathtt{KR}} = \left[x_{t,0}^{\mathtt{KR}}\ x_{t,1}^{\mathtt{KR}}\ \cdots\ x_{t,10}^{\mathtt{KR}}\right]^{\mathsf{T}}$. The *autoregressive* component $x_t^{\mathtt{AR}}$ models the time-dependent residual term representing the effect of other phenomena that cannot be modeled by the components employed.

Moreover, a BDLM involves a set of parameters $\mathcal{P}$ that is learned from the observed data.
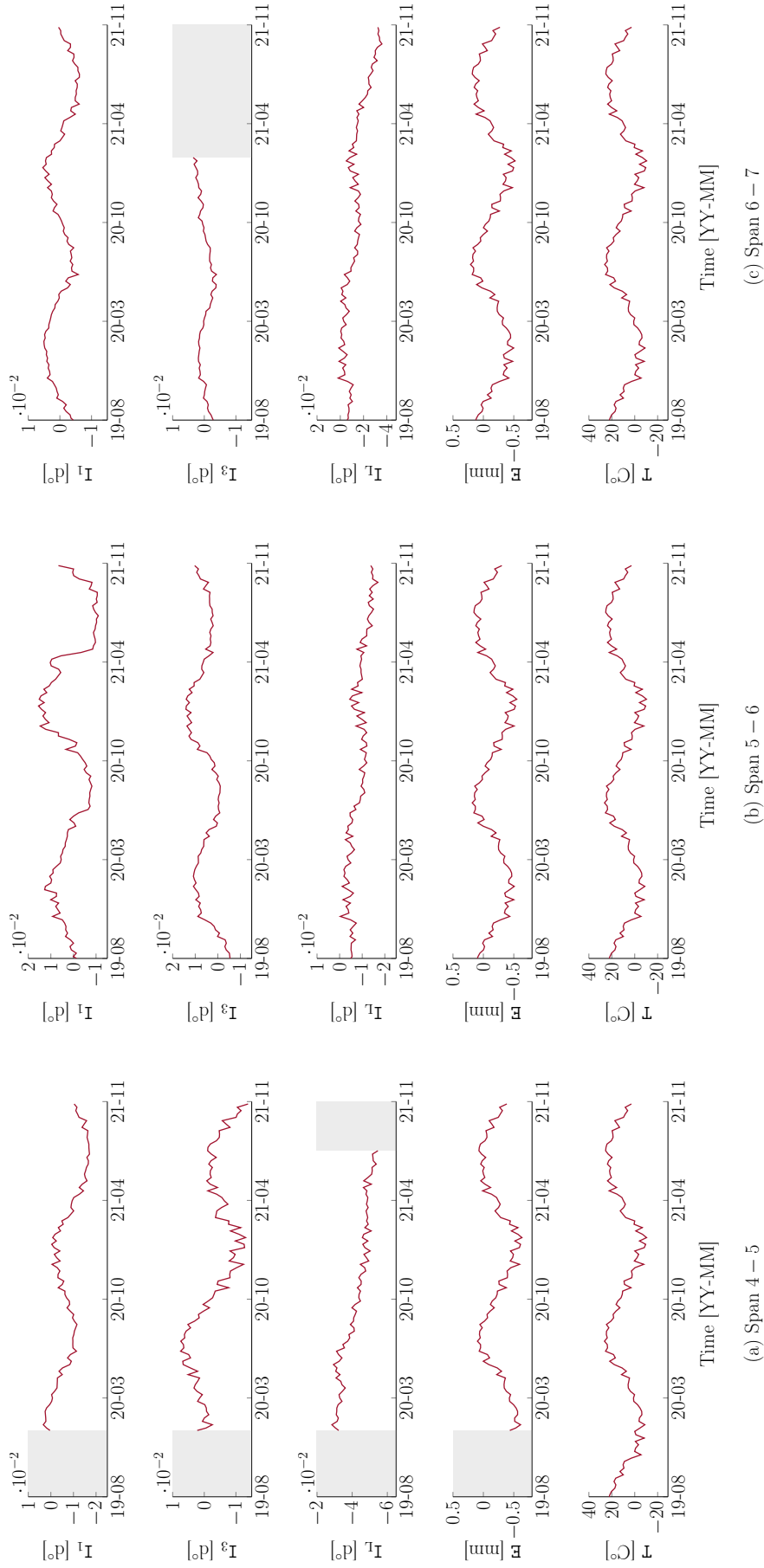
Figure 3.7 Illustration of the preprocessed weekly data (a) Span 4 − 5, (b) Span 5 − 6, and (c) Span 6 − 7. The gray areas represent either missing or improperly collected data.

In general, the parameter set consist in $\sigma_w \in \mathbb{R}^+$, the standard deviation of the process noise for each irreversible component; $\ell \in \mathbb{R}^+$ and $\sigma \in \mathbb{R}^+$ are the kernel length and the standard deviation of the hidden periodic pattern corresponding to the kernel regression component; and $\phi^{\mathtt{AR}} \in (0, 1]$, the autoregressive coefficient. Here, we assume a fixed kernel period equal to 365.2422 days and a zero standard deviation for the hidden control points. Apart from the parameters corresponding to the components, there are additional ones that need to be learned; They are $\sigma_v \in \mathbb{R}^+$, the observation standard deviation for each measurement, and regression coefficients β expressing the dependencies between time series data. Regression coefficients are introduced in the observation model matrix $\mathbf{C}_t$ in order to create a linear dependency between each measurement and temperature time series. Appendix A provides the observation model matrices for modeling time series.

Table 3.2 summarizes the list of generic components used to model the time series obtained from the *extensometer, inclinometer*, and *thermometer*, the three types of sensors used in this thesis. The first column identifies the name of each component followed by its schematic symbol and mathematical notation, and the last column provides the corresponding parameters that either are fixed or need to be learned from data. The hidden state variables estimation and the optimization of model parameters $\mathcal{P}^*$ are carried out by using the openBDLM library [84].

Table 3.2 BDLM generic components used to model a non-stationary regime. The columns show the names, symbols, notations, and parameters for each hidden state variable. Note that parameters with fixed values are not optimized. The icons are adopted and edited from [1].

Component	Icon	Notation	Parameter
Level		$x^{\mathtt{L}}$	$\sigma_w^{\mathtt{L}} = 0$
Trend		$x^{\mathtt{T}}$	$\sigma_w^{\mathtt{T}} = 0$
Local acceleration (acceleration)		$x^{\mathtt{LA}}$	$\sigma_w^{\mathtt{LA}} = 0$
Kernel regression		$\mathbf{x}^{\mathtt{KR}}$	$\ell,\ \sigma_w^{\mathtt{KR}} = 0$
Autoregressive		$x^{\mathtt{AR}}$	$\phi^{\mathtt{AR}},\ \sigma_w^{\mathtt{AR}}$
Observation		$\mathbf{y}$	

The only observed external effect is the air temperature. Therefore, it is required to establish dependencies between the periodic and autoregressive components of the temperature data and other measurements. Figure 3.8 shows the schematic of the generic BDLM model for non-stationary regimes $\mathcal{M}^{\text{ns}}$ corresponding to inclinometers (Figure 3.8a) and extensometers (Figure 3.8b). The diagrams represent the components selected in order to decompose the

(a) Inclinometer
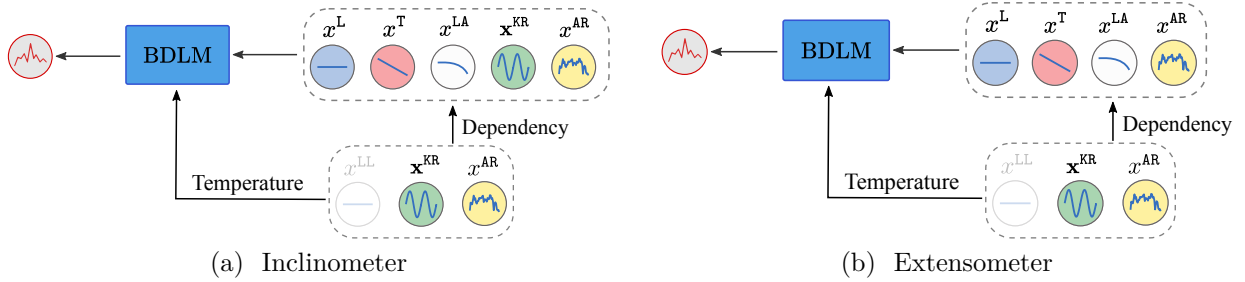
(b) Extensometer

Figure 3.8 Illustration of the schematics for the generic BDLM models associated with the non-stationary regime across three spans: (a) inclinometer, and (b) extensometer. The icons used is in accordance with Table 3.2. A stationary regime follows the same diagrams except that it does not include the acceleration. The only dependency is between the structural responses and the temperature data. Such a dependency is established throughout the kernel regression and autoregressive components.

time series into reversible and irreversible responses. Note that the generic BDLM model for the stationary regime $\mathcal{M}^{\text{s}}$ remains the same as those presented in Figure 3.8 except that it does not include the acceleration. Therefore, the irreversible responses consist in a level x^{L} and a local trend x^{LT} hidden state variables. The remaining part of this section provides the results obtained using BDLMs.

3.5.2 Thermometer

The vector of hidden state variables $\mathbf{x}_t$ for each empirical model of temperature data follows

$$\mathbf{x}_t = \left[x_t^{\text{LL,T}}; \mathbf{x}_t^{\text{KR,T}}; x_t^{\text{AR,T}} \right], \tag{3.1}$$

where $[\,;]$ indicates the column-wise concatenation of the vectors and the additional superscript $^{\text{T}}$ indicates that the observations are for the temperature. Note that building the empirical model for temperature data only requires the stationary model $\mathcal{M}^{\text{s}}$ because we do not expect to have a regime switch. Appendix A provides the detailed model matrices set

$\mathcal{M}^{\mathrm{s}}$. Moreover, the model parameters that learned from data are

$$\mathcal{P} = \left\{ \ell^{\mathrm{KR,T}}, \phi^{\mathrm{AR,T}}, \sigma^{\mathrm{AR,T}}, \sigma_v^{\mathrm{T}} \right\}. \tag{3.2}$$

Table 3.3 shows the optimized parameters set $\mathcal{P}^*$ for all three spans. The first column shows the name of the spans, and the remaining columns are the values corresponding to the estimated parameters which their values are similar; This indicates that the temperature data decomposition results in similar hidden state variable estimations.

Figure 3.9 illustrates the time series decomposition associated with the temperature measurements. We use the same model matrices in order to build the empirical models for all
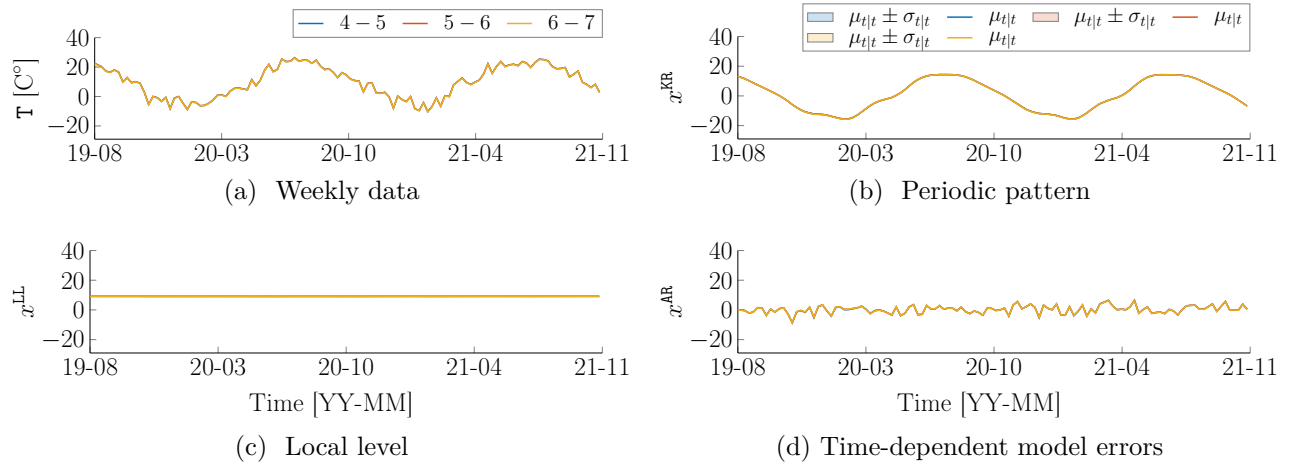
Figure 3.9 The hidden state estimations for temperature data across all instrumented spans $4 - 5$, $5 - 6$, and $6 - 7$. The solid lines are associated with the expected values $\mu_{t|t}$ for each hidden state variable, and the shaded areas represent the uncertainty $\mu_{t|t} \pm \sigma_{t|t}$ around these expected values. Note that the uncertainty around the estimations is small that is not visible.

three temperature measurements across all three spans. The generic components involve the local level, kernel regression, and autoregressive components. As it is seen, the hidden state estimations for all three spans are coinciding.

The temperature data affect other measurements collected from the inclinometers and extensometers. Such effects are modeled by creating dependencies between the kernel regression and autoregressive hidden state variables of temperature measurements with the ones for other sensors. In the next section, we introduce two coefficients β^{KR} and β^{AR} in the observation matrix $\mathbf{C}_t$ in order to account for such dependencies.

Table 3.3 Optimized parameters set $\mathcal{P}^*$ for the BDLM models associated with the temperature data across all three spans.

Span	$\ell^{\text{KR,T}}$	$\phi^{\text{AR,T}}$	$\sigma^{\text{AR,T}}$	σ_v^{T}
$4-5$	0.46	0.07	2.61	0.10
$5-6$	0.46	0.08	2.65	0.10
$6-7$	0.45	0.07	2.64	0.10

3.5.3 Inclinometer

We use an identical model structure for all inclinometer data as shown in Figure 3.8a. The vector of hidden state variables $\mathbf{x}_t$ for each empirical model follows

$$\mathbf{x}_t = \left[\underbrace{\mathbf{x}_t^{\text{LA,I}}; \mathbf{x}_t^{\text{KR,I}}; x_t^{\text{AR,I}}}_{\text{inclination components}} \; ; \; \underbrace{x_t^{\text{LL,T}}; \mathbf{x}_t^{\text{KR,T}}; x_t^{\text{AR,T}}}_{\text{temperature components}} \right], \tag{3.3}$$

where the additional superscript $^{\text{I}}$ indicates that the observations are related to the inclinometer data. In Equation 3.3, the local acceleration vector $\mathbf{x}_t^{\text{LA}} = [x_t^{\text{L}} \; x_t^{\text{T}} \; x_t^{\text{LA}}]^{\mathsf{T}}$ describes the non-stationary regime of the dynamic system as presented in Equation 2.4. For the stationary regime, the local acceleration is equal to zero for all time steps following the transition matrix and error presented in Equation 2.6. Appendix A provides the detailed model matrices for both regimes, i.e., $\mathcal{M}^{\text{s}}$ and $\mathcal{M}^{\text{ns}}$. Furthermore, we use the same model parameters presented in Table 3.3 for temperature data. Therefore, the only parameters that need to be learned from data are associated with inclinations, transition probabilities, and the dependency coefficients following

$$\mathcal{P} = \left\{ \underbrace{z^{\text{s,s}}, z^{\text{ns,ns}}}_{\text{regime transition}}, \underbrace{\sigma_{\text{Tr}}^{\text{LA,I}}, \ell^{\text{KR,I}}, \phi^{\text{AR,I}}, \sigma^{\text{AR,I}}, \sigma_v^{\text{I}}}_{\text{inclination, I}}, \underbrace{\beta^{\text{AR,I|T}}, \beta^{\text{KR,I|T}}}_{\text{dependency, I|T}}, \right\}, \tag{3.4}$$

where $z^{i,i}$ is the transition probability between the regimes, and $\sigma_{\text{Tr}}^{\text{LA}} \in \mathbb{R}^+$ is the standard deviation of the transition model that corresponds to the transition from a stationary to a non-stationary regime. In Equation 3.4, two dependencies are introduced. They correspond to the dependency between the temperature and inclination data with respect to the autoregressive component, i.e., $\beta^{\text{AR,I|T}}$, and the kernel regression component $\beta^{\text{KR,I|T}}$. These two coefficients determine how temperature data affect the structural measurements. In what follows, we present the results for the data collected from the lateral and longitudinal inclinometers.

Lateral Inclinometer

The lateral inclinometer for each span is located on the downstream side of the axis 3 as shown in Figure 3.3. The regime transition probabilities for all three spans are fixed such that $z^{\text{s,s}} = 0.9999$ and $z^{\text{ns,ns}} = 0.9995$. In addition, for these spans, the observation standard deviation is fixed to $\sigma_v^{\text{I}} = 0.0001$ and the transition model standard deviation is fixed to $\sigma_{\text{Tr}}^{\text{LA}} = 1 \times 10^{-7}$. Table 3.4 shows the optimized parameter values. The first column identifies the span names

Table 3.4 Optimized parameters for the BDLM models associated with the lateral inclinometer across all three spans. The columns are divided into two categories with respect to the inclination data, and the dependency coefficients.

Span	Inclination, I			Dependency ($\times 10^{-4}$)	
	$\ell^{\text{KR,I}}$	$\phi^{\text{AR,I}}$	$\sigma^{\text{AR,I}}$	$\beta^{\text{AR,I|T}}$	$\beta^{\text{KR,I|T}}$
$4-5$	0.86	0.11	0.001	-2.12	-2.96
$5-6$	5.21	0.55	0.001	-4.40	-1.39
$6-7$	3.85	0.58	0.002	-6.33	-2.89

and the following three columns show the optimized parameter values associated with the inclination data. The last two columns are the dependency values between the temperature and inclination data. The dependency coefficients confirm that there is a negative linear dependency between the lateral inclinometer data and temperature data.

Figure 3.10 illustrates the hidden state estimations along with the probability of the non-stationary regimes for all instrumented spans. As it is seen, the span $4-5$ hidden state estimations differ from the other two spans; In Figure 3.10c, the level response x^{L} shows a shift indicating that the lateral inclinometer suggests a different degradation for the span $4-5$ compared with the other ones. Such a difference is also seen for the trend responses x^{T} as depicted in Figure 3.10e. Figure 3.10b shows the periodic patterns that are not displayed by the temperature data. For the lateral inclinometer on the span $4-5$, it shows that such a periodic external effect is different from the other two spans. This is in agreement with the values obtained for the kernel regression length $\ell^{\text{KR,I}}$ as presented in the second column of Table 3.4 in the sense that different kernel lengths result in different patterns. In Figure 3.10f, the probabilities indicate regime switches around October 2020 for the spans $5-6$ and $6-7$. For the span $4-5$, a regime switch with a probability of approximately 0.5 happened around March 2021. We are not aware of the cause of the aforementioned regime switches; since they happened on three spans approximately around the same time, we do not presume they are structural anomalies. Rather, we suspect that they are due to intervention, sensor malfunctioning, and other non-structural damages.

(a) Weekly data

(b) Periodic pattern no-observed by the temperature

(c) Level

(d) Time-dependent model errors

(e) Trend

(f) Probability of the non-stationary regime

(g) Local acceleration
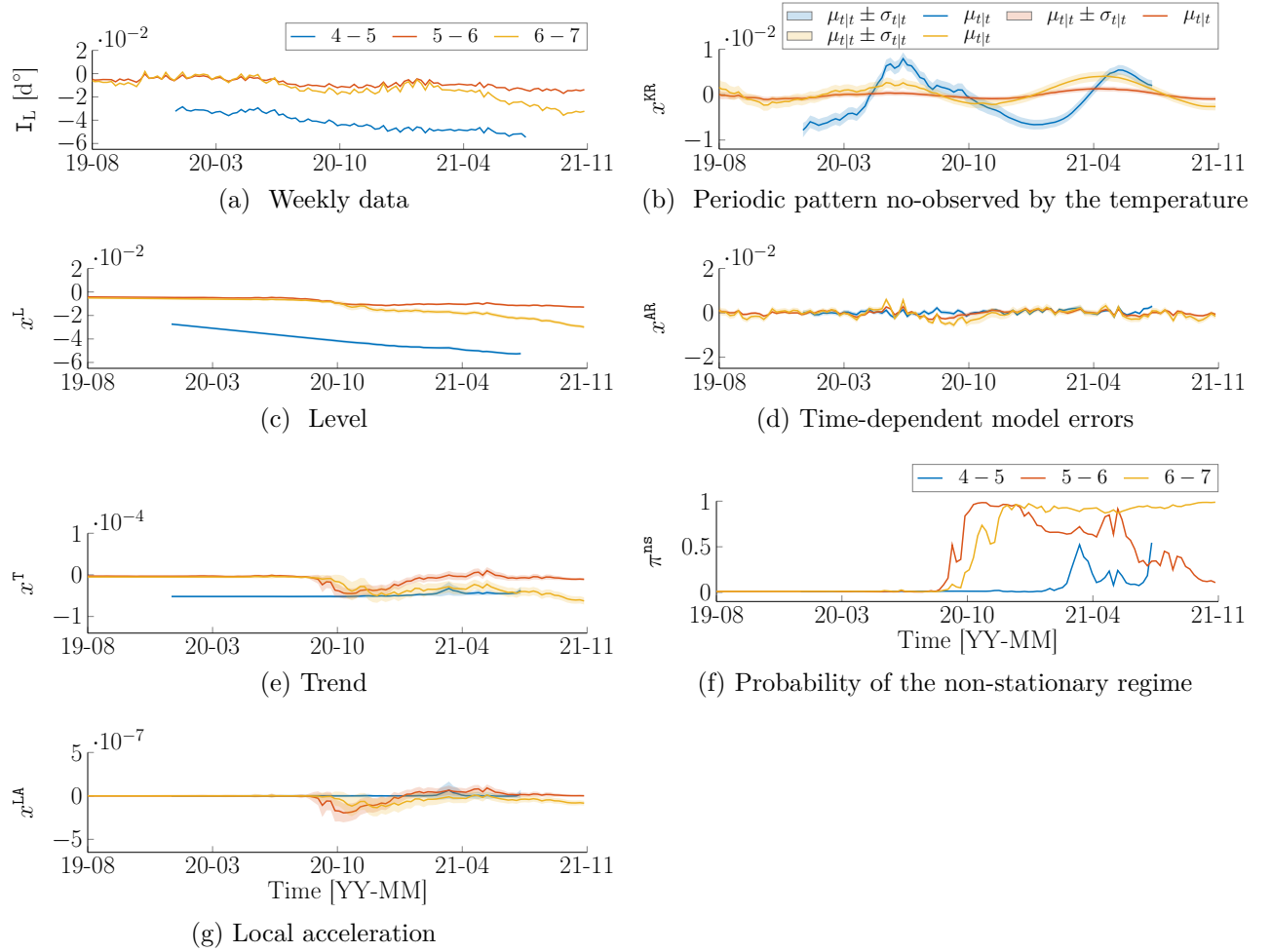
Figure 3.10 Illustration of the hidden state estimations and the probability of the non stationary regimes for the lateral inclination data on the axis 3 across all instrumented spans $4-5$, $5-6$, and $6-7$. The solid lines are associated with the expected values $\mu_{t|t}$ for each hidden state variable, and the shaded areas represent the uncertainty $\mu_{t|t} \pm \sigma_{t|t}$ around these expected values.

Longitudinal Inclinometer 1

The longitudinal inclinometer 1 for each span is located on the downstream side of the axis 1 as shown in Figure 3.3. The regime transition probabilities for all three spans are fixed such that $z^{\text{s,s}} = 0.9999$ and $z^{\text{ns,ns}} = 0.9990$. In addition, for these spans, the observation standard deviation is set to $\sigma_v^{\text{I}} = 0.0001$ and the transition model standard deviation is fixed to $\sigma_{\text{Tr}}^{\text{LA}} = 1 \times 10^{-7}$. Table 3.5 shows the optimized parameter values. Similarly to the

Table 3.5 Optimized parameters for the BDLM models associated with the longitudinal inclinometer 1 across all three spans. The columns are divided into two categories with respect to the inclination data and the dependency coefficients.

Span	Inclination, I			Dependency ($\times 10^{-4}$)	
	$\ell^{\text{KR,I}}$	$\phi^{\text{AR,I}}$	$\sigma^{\text{AR,I}}$	$\beta^{\text{AR,I|T}}$	$\beta^{\text{KR,I|T}}$
$4-5$	1.67	0.68	0.0004	-3.51	-3.77
$5-6$	4.42	0.43	0.0009	-3.64	-4.16
$6-7$	0.81	0.52	0.0005	-1.56	-2.63

lateral inclinometer data, the last two columns in Table 3.5 indicate that the longitudinal inclinometer 1 has a linear negative dependency with temperature data.

Figure 3.11 illustrates the hidden state estimations along with the probability of the non-stationary regimes. In Figure 3.11c, there is a shift in the level x^{L} corresponding to the span $4-5$ compared with the other two spans. In addition, the periodic component for the span $6-7$ is different compared with two other spans as shown in Figure 3.11b; Figure 3.11f illustrates the probability of the non-stationary regimes for the three spans. The span $5-6$ shows two regime switches with a probability of 1 during August 2020 and May 2021. The span $4-5$ shows a regime switch that started in April 2021 with a probability of around 0.3 which reaches 1 around July 2021. However, the analysis of the longitudinal inclinometer data for the span $6-7$ indicates that there was no anomaly during the time from 2019 to 2021.

Longitudinal Inclinometer 3

The longitudinal inclinometer 3 for each span is located on the downstream side of the axis 3 as shown in Figure 3.3. The regime transition probabilities for all three spans are fixed such that $z^{\text{s,s}} = 0.9999$ and $z^{\text{ns,ns}} = 0.9995$. In addition, for these spans, the observation standard deviation is fixed to $\sigma_v^{\text{I}} = 0.0001$ and the transition model standard deviation is set to $\sigma_{\text{Tr}}^{\text{LA}} = 1 \times 10^{-7}$. Table 3.6 shows the optimized parameter values. Similarly to the
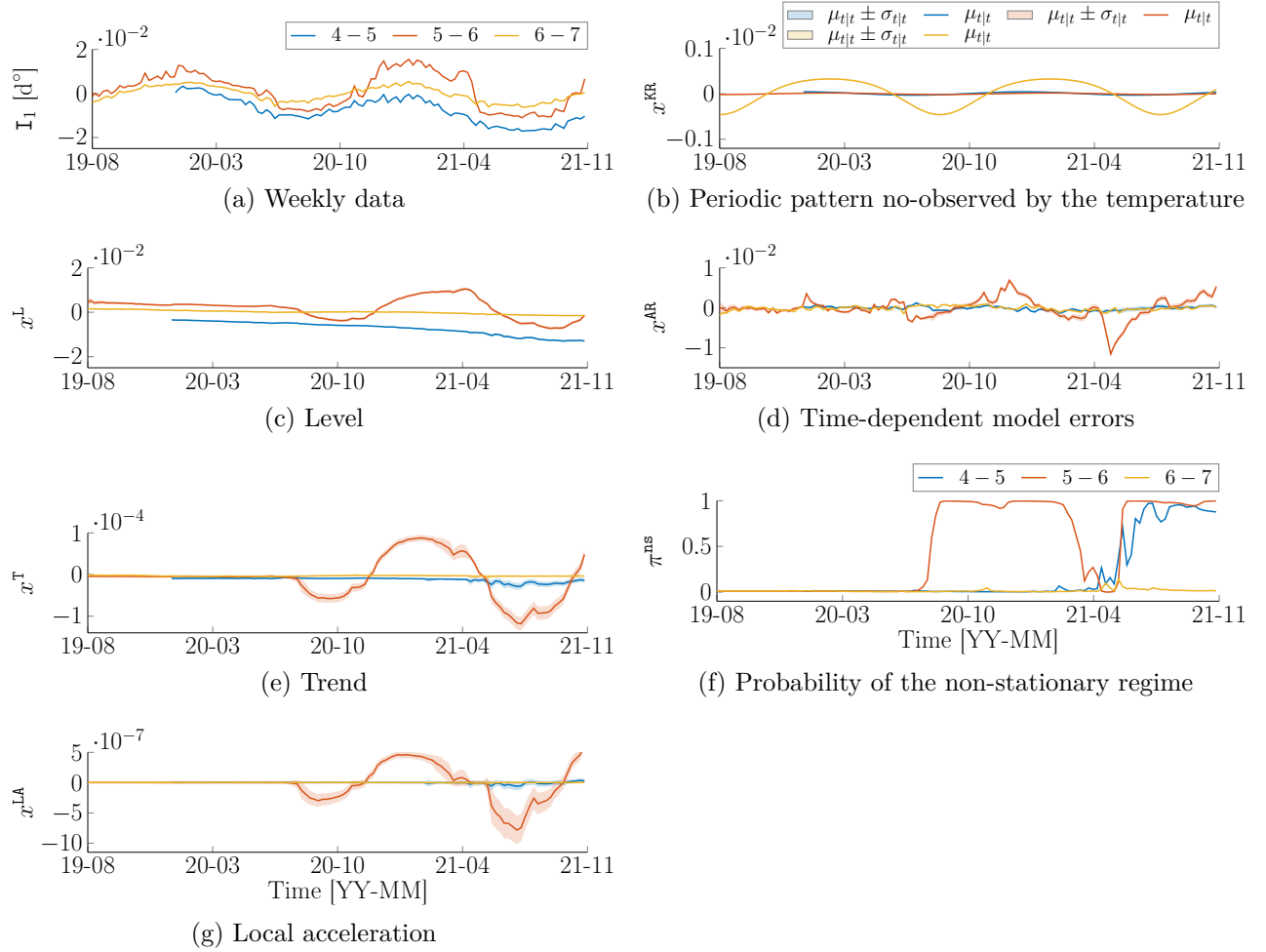
Figure 3.11 Illustration of the hidden state estimations and the probability of the non-stationary regimes for the longitudinal inclination data on the axis 1 across all instrumented spans $4-5$, $5-6$, and $6-7$. The solid lines are associated with the expected values $\mu_{t|t}$ for each hidden state variable, and the shaded areas represent the uncertainty $\mu_{t|t} \pm \sigma_{t|t}$ around these expected values.

lateral and longitudinal inclinometers data, the last two columns in Table 3.6 indicate that the longitudinal inclinometer 3 data has a linear negative dependency with temperature data, but only for the spans $5-6$ and $6-7$. In contrast, the dependency coefficients for the span $4-5$ is positive; we suspect that this sensor is installed in a reverse direction resulting in such positive dependencies.

Figure 3.12 illustrates the hidden state estimations along with the probability of the non-stationary regimes obtained for the longitudinal inclinometers 3 across all three instrumented spans. From Figure 3.12a, we suspect that the measurements for the span $4-5$ is collected differently compared with the other two spans; This is confirmed by the dependency coeffi-

Table 3.6 Optimized parameters for the BDLM models associated with the longitudinal inclinometer 3 across all three spans. The columns are divided into two categories with respect to the inclination data and the dependency coefficients.

Span	Inclination, $\mathtt{I}$			Dependency $(\times 10^{-4})$	
	$\ell^{\mathtt{KR,I}}$	$\phi^{\mathtt{AR,I}}$	$\sigma^{\mathtt{AR,I}}$	$\beta^{\mathtt{AR,I\mid T}}$	$\beta^{\mathtt{KR,I\mid T}}$
$4-5$	2.51	0.48	0.0004	$+5.11$	$+3.28$
$5-6$	0.36	0.67	0.0003	-3.06	-4.42
$6-7$	4.57	0.74	0.0004	-1.00	-1.64

(a) Weekly data

(b) Periodic pattern no-observed by the temperature

(c) Level

(d) Time-dependent model errors

(e) Trend

(f) Probability of the non-stationary regime

(g) Local acceleration
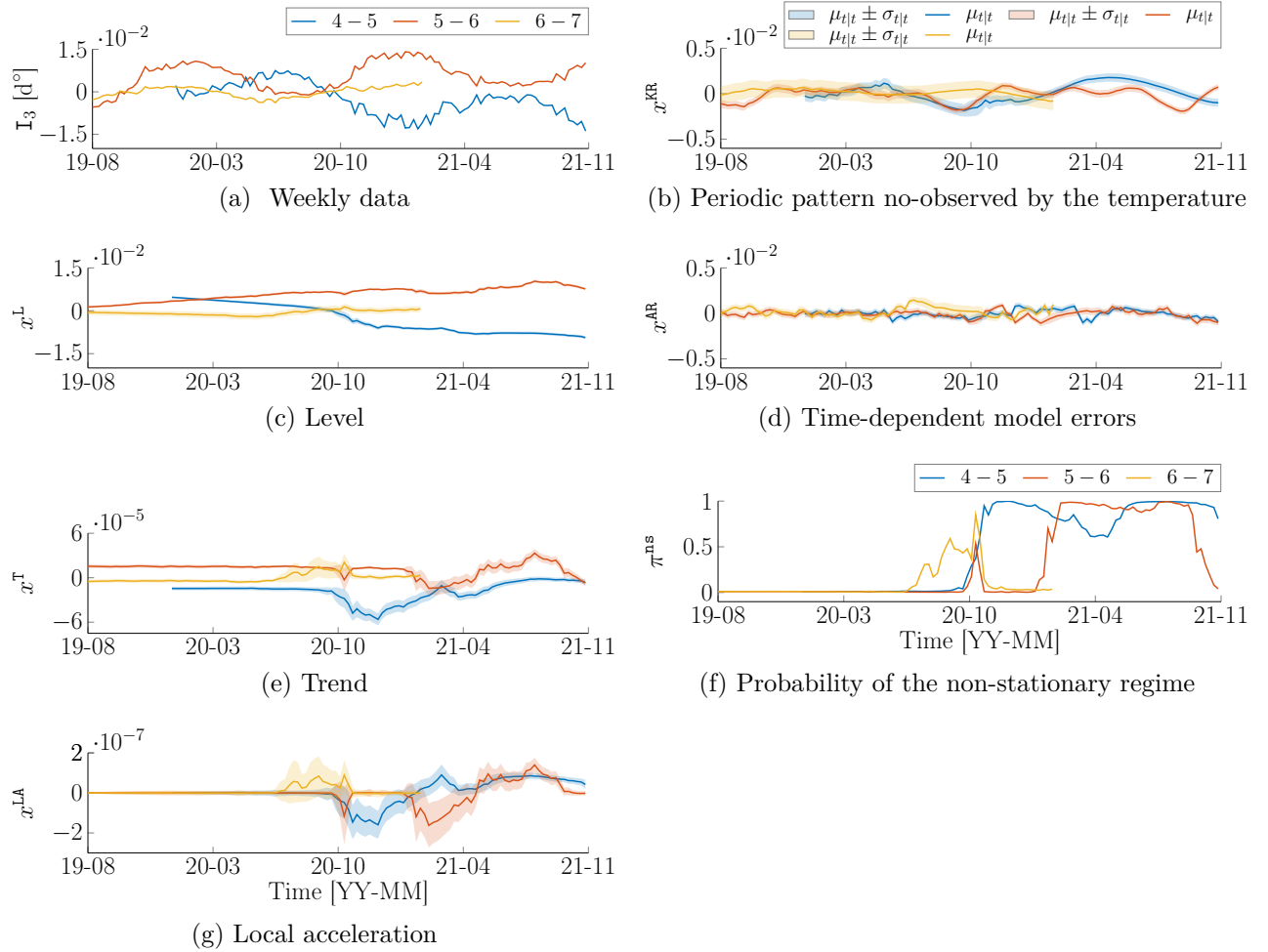
Figure 3.12 Illustration of the hidden state estimations and the probability of the non-stationary regimes for the longitudinal inclination data on the axis 3 across all instrumented spans $4-5$, $5-6$, and $6-7$. The solid lines are associated with the expected values $\mu_{t|t}$ for each hidden state variable, and the shaded areas represent the uncertainty $\mu_{t|t} \pm \sigma_{t|t}$ around these expected values.

cients presented in Table 3.6 for which the coefficients for the span $4 - 5$ have the opposite sign compared with those associated with other spans which have a negative sign. In addition, Figure 3.12e shows the trend response for the measurement data. The span $5 - 6$ shows a positive degradation's speed for the stationary regime, while it is negative for the spans $4 - 5$ and $6 - 7$. Another point that needs to be mentioned is related to the kernel regression hidden state shown in Figure 3.12b; The periodic component for the span $5 - 6$ differs from the other two spans, which is confirmed by different values obtained for the kernel regression lengths $\ell^{\mathtt{KR,I}}$ presented in Table 3.6. Figure 3.12f illustrates the probability of the non-stationary regimes for the instrumented spans. The spans $4 - 5$ and $6 - 7$ indicate regime switches with a probability close to one around October 2020. The span $5 - 6$ shows two regime switches with probabilities of 0.6 and around October 2020 and February 2021.

3.5.4 Extensometer

We use the same generic model for all extensometer data as shown in Figure 3.8b. The vector of hidden state variables $\mathbf{x}_t$ for each empirical model follows

$$\mathbf{x}_t = \left[\underbrace{\mathbf{x}_t^{\mathtt{LA,E}}; x_t^{\mathtt{AR,E}}}_{\text{elongation components}} \quad ; \quad \underbrace{x_t^{\mathtt{LL,T}}; \mathbf{x}_t^{\mathtt{KR,T}}; x_t^{\mathtt{AR,T}}}_{\text{temperature components}} \right], \tag{3.5}$$

where the additional superscript $^{\mathtt{E}}$ indicates that the observations are for the elongation data collected from extensometers. Analogously to the inclinometer empirical models, the local acceleration vector $\mathbf{x}_t^{\mathtt{LA}} = [x_t^{\mathtt{L}}, x_t^{\mathtt{T}}, x_t^{\mathtt{LA}}]^{\mathtt{T}}$ in Equation 3.5 describes the non-stationary regime of the dynamic system as presented Equation 2.4. For the stationary regime, the local acceleration is equal to zero for all time steps following the transition matrix and error presented in Equation 2.6. Appendix A provides the detailed model matrices for both regimes, i.e., $\mathcal{M}^{\mathtt{s}}$ and $\mathcal{M}^{\mathtt{ns}}$. Again, we use the same model parameters presented in Table 3.3 for temperature data. Therefore, the only parameters that need to be learned from the data are associated with the elongation, the transition probabilities, and the dependency coefficients following

$$\mathcal{P} = \left\{ \underbrace{z^{\mathtt{s,s}}, z^{\mathtt{ns,ns}}}_{\text{regime transition}}, \underbrace{\sigma_{\mathrm{Tr}}^{\mathtt{LA,E}}, \phi^{\mathtt{AR,E}}, \sigma^{\mathtt{AR,E}}, \sigma_v^{\mathtt{E}}}_{\text{elongation, E}}, \underbrace{\beta^{\mathtt{AR,I|T}}, \beta^{\mathtt{KR,E|T}}}_{\text{dependency, E|T}} \right\}, \tag{3.6}$$

where, two dependencies similar to the ones in Equation 3.4 are introduced for the extensometer data. They correspond to the dependency between the temperature and elongation with respect to the autoregressive $\beta^{\mathtt{AR,E|T}}$ and the kernel regression $\beta^{\mathtt{KR,E|T}}$ components. The regime

transition probabilities for all three spans are fixed such that $z^{\text{s,s}} = 0.9999$ and $z^{\text{ns,ns}} = 0.9990$. Moreover, the observation standard deviation is fixed to $\sigma_v^{\text{E}} = 0.001$ and the transition model standard deviation is set to $\sigma_{\text{Tr}}^{\text{LA}} = 1 \times 10^{-7}$. Table 3.7 shows the optimized parameter values. The last two columns indicate that the elongation data have a linear positive dependency with respect to the temperature. Figure 3.13 illustrates the hidden state estimations along

Table 3.7 Optimized parameter values for the BDLM models associated with the extensometers across all three spans. The columns are divided into two categories of optimized parameters associated with the elongation data and the dependency coefficients.

Span	Inclination, E		Dependency	
	$\phi^{\text{AR,E}}$	$\sigma^{\text{AR,I}}$	$\beta^{\text{AR,E|T}}$	$\beta^{\text{KR,E|T}}$
$4-5$	0.82	0.004	0.02	0.02
$5-6$	0.82	0.003	0.02	0.02
$6-7$	0.80	0.003	0.02	0.02

with the probability of the non-stationary regimes for all instrumented spans. The elongation data corresponding to the span $4-5$ has a shift shown in Figure 3.13c. The probability of the non-stationary regimes indicate that the regime is stationary for all the time steps during 2019 and 2021.

52
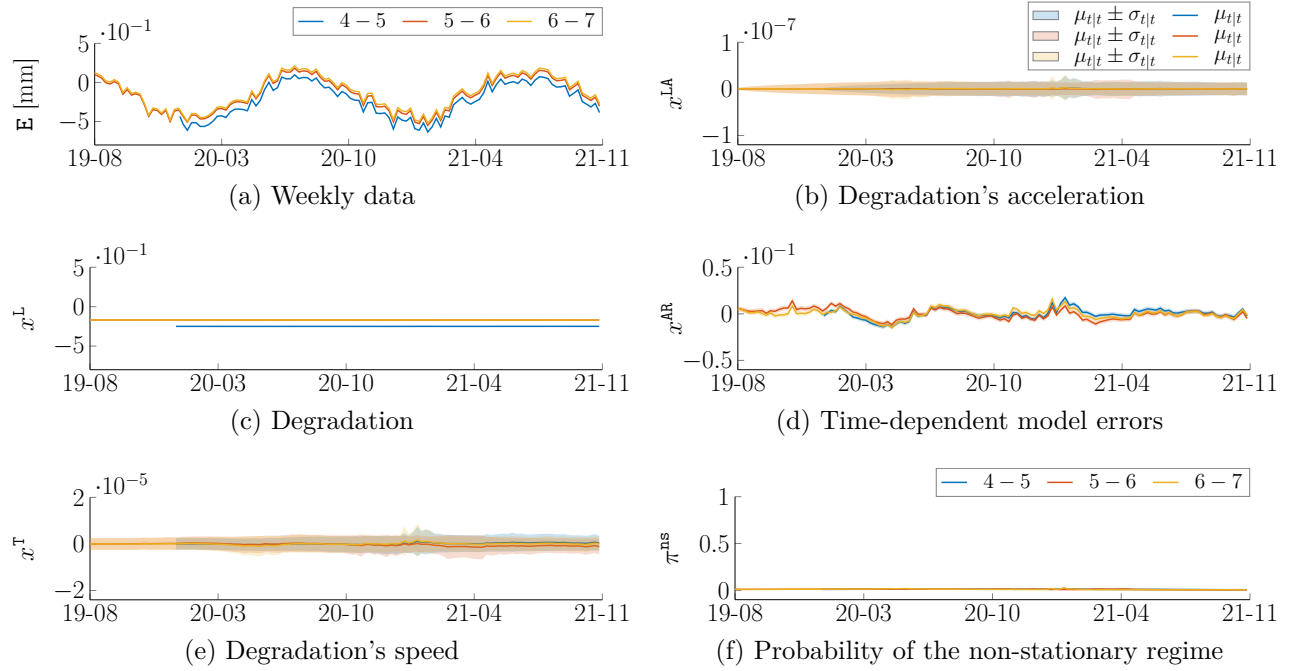
Figure 3.13 Illustration of the hidden state estimations and the probability of the non-stationary regimes for the elongation data across all instrumented spans $4-5$, $5-6$, and $6-7$. The solid lines are associated with the expected values $\mu_{t|t}$ for each hidden state variable, and the shaded areas represent the uncertainty $\mu_{t|t} \pm \sigma_{t|t}$ around these expected values.

3.6   Discussion

We identified many issues along with the regime switches associated with the inclinometers. In order to evaluate the performance of these sensors, we simulated synthetic time series using their respective empirical models associated with a stationary regime as shown in Figure 3.14. The subscript 1:15 indicates that the mean and standard deviation are computed for 15 simulated time series. For the thermometers and extensometers, the synthetic time series coincide with the real measurements in the sense that they follow the same patterns. However, for the inclinometers, there are some cases that there is a match between the real and synthetic time series, e.g., the longitudinal inclinometer 1 on span $4-5$ and $6-7$. Nevertheless, in many cases there are drifts between the simulated and real time series, and we do not see common tendencies in the drift patterns. In addition, in some cases such drifts are reversed; For instance, the elongations for the longitudinal 3 show a downward trend for the span $4-5$, while it is upward for the span $6-7$.

Furthermore, in some cases we see interruptions or jumps in data associated with the longitudinal inclinometer 1 on span $5-6$ and the longitudinal inclinometer 3 on span $6-7$.
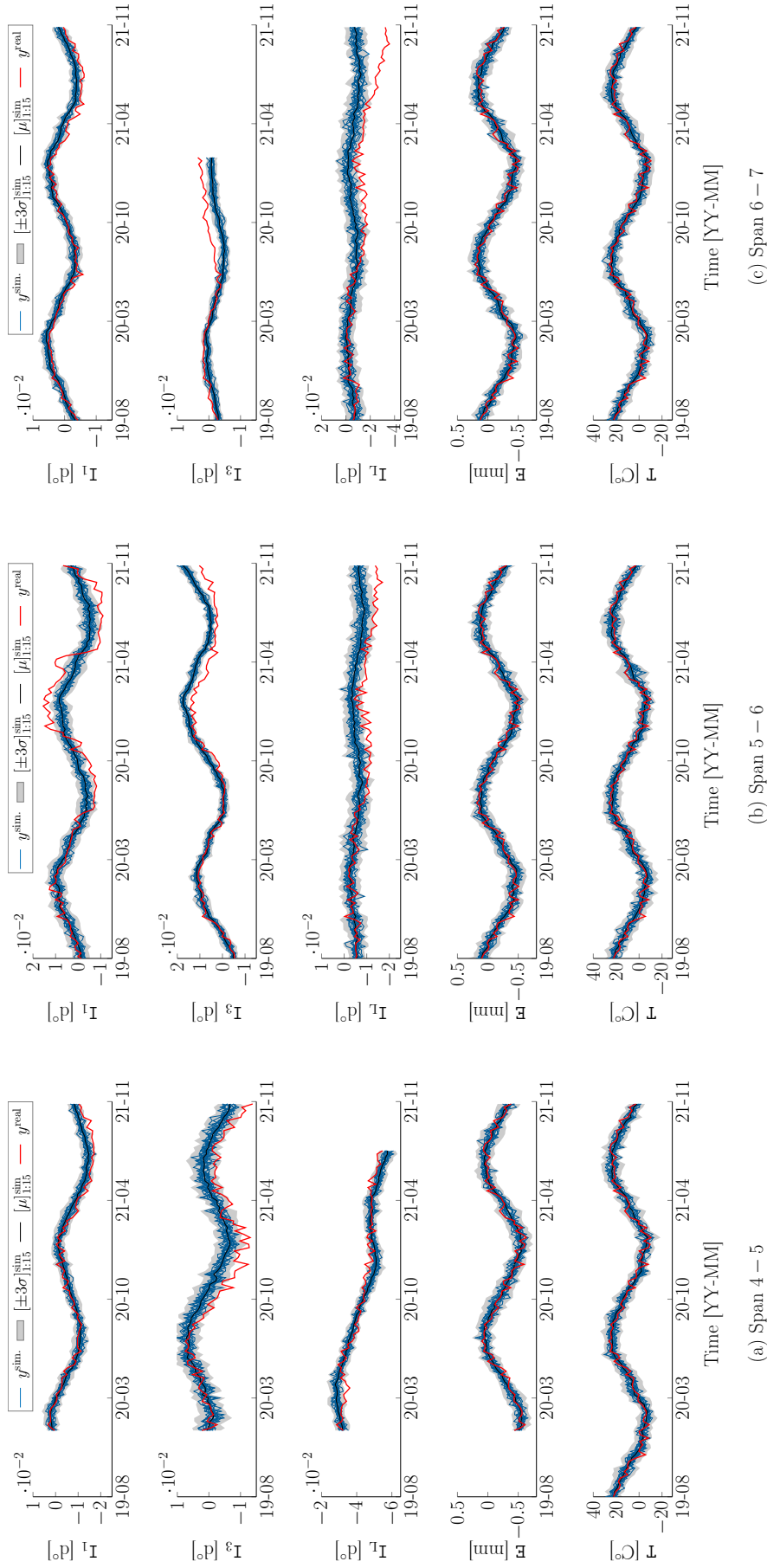
Figure 3.14 Illustration of simulating 15 synthetic time series using stationary model $\mathcal{M}^{\mathbf{s}}$: (a) Span $4-5$, (b) Span $5-6$, and (c) Span $6-7$. The solid blue lines are the simulated data y^{sim}, while the red ones are the real data y^{real} collected from the respective sensors. The solid black lines are the expected values $[\mu]_{1:15}^{\text{sim}}$ of the simulated time series and the shaded areas represent their three standard deviation $[\pm 3\sigma]_{1:15}^{\text{sim}}$ around the simulated data.

The former inclinometer shows a jump in the collected data during October 2021 and April 2022. The latter shows that the measurements are not collected after February 2021. From the above-mentioned issues, we conclude that the main issue is with the inclinometers themselves. As a result, the repeated interruptions, drifts, and jumps in the collected data make us believe that these sensors are not adequate for the long-term monitoring of the Jacques Cartier Bridge because they are unable to have a long-term stability in the acquisition of structural responses. However, for the purpose of validating the methodologies developed in this thesis, we will use the empirical models built for the inclinometers.

3.7   Conclusion

This chapter presented the analysis of the data collected on the Jacques Cartier Bridge. We employed BDLM in order to analyze the measurements associated with the inclinometers, extensometers, and thermometers. The analysis of the measurements led to the decomposition of the data into reversible and irreversible responses for each sensor along with the probability of the non-stationary regimes for inclinometers and extensometers. The BDLM models identified regime switches during the time steps between August 2019 and November 2022. In addition, we identified many issues regarding interruptions, drifts, and jumps in data associated with the inclinometers. These issues makes us believed that choosing the inclinometers is not suited for the long-term monitoring of the Jacques Cartier Bridge. However, we will use the empirical models built for inclinometers data for the purpose of validating the methodologies developed in this thesis. In contrast, the empirical models for the thermometers and the extensometers showed a stable acquisition of temperature and elongation data that is suited for the long-term monitoring of the bridge. The analysis of the time series resulted in empirical models for each sensor. Such models can be used in order to generate new time series mimicking the structural responses. The next two chapters will present methodologies based on the reinforcement learning and imitation learning in order to leverage the BDLM empirical models for developing an anomaly detection framework. This framework benefits from (i) the interpretability of the BDLM models in order to extract irreversible responses subject to a wide range of anomalies, and (ii) incorporating the information obtained from the BDLM analysis consisting in the structural degradation's speed and the probability of the non-stationary regime for decision making.

CHAPTER 4   Anomaly Detectability Using Reinforcement Learning

This chapter presents the methodologies developed for detecting anomalies and for quantifying the detectability of anomalies, along with the application of these methods to the Jacques Cartier Bridge. We first present the components behind the anomaly detection method based on the reinforcement learning (RL) approach: (i) an environment based on a decomposition of time series using Bayesian dynamic linear models (BDLM), (ii) a procedure for generating synthetic time series representing the structural responses subject to a wide range of anomalies, and (iii) a reward function based on true and false detections in order to encourage or prevent an agent to take actions at different time steps. Moreover, we present how we propose to quantify the anomaly detectability with respect to the anomaly's magnitude, its duration, and the time required to detect an anomaly after its starting time. Such a quantification leads to two probabilistic metrics; The *annual false detection ratio*, and the *probability of true positive detection*, which characterize probabilistically the capacity of each sensor in detecting anomalies. Finally, we validate these methodologies on the measurements collected on the three instrumented spans of the Jacques Cartier Bridge.

4.1   Anomaly Detection Using Reinforcement Learning

Making decisions regarding the presence of an anomaly requires incorporating the information from both the irreversible responses of a structure and the probability of being in a non-stationary regime. In this research, these two quantities are represented by the expected value of the degradation's speed $\mu_t^{\mathtt{T}}$ known as the *trend* ($\mathtt{T}$), and the probability of a non-stationary ($\mathtt{ns}$) regime $\pi_t^{\mathtt{ns}}$. They jointly form the environment's state vector $\mathbf{s}_t = [\mu_t^{\mathtt{T}} \ \pi_t^{\mathtt{ns}}]^{\mathsf{T}} \in \mathcal{S}$.

Figure 4.1 shows an example of a continuous *environment's state-space* $\mathcal{S}$ for which the solid and dashed lines represent the evolution of a vector $\mathbf{s}_t$ over time. An *episode* $\mathcal{Z} = \{\mathbf{s}_1, \mathbf{s}_2, \cdots\}$ is defined as a set of environment's state vectors for consecutive time steps. Episodes such as the one depicted by the solid line are initialized at $\mathbf{s}_1$, at which time it follows a stationary regime with a zero degradation's speed. An anomaly starts at the time $t = t_{\mathrm{s}}$ which corresponds to the state $\mathbf{s}_{t_{\mathrm{s}}}$. At a time $t > t_{\mathrm{s}}$ after an anomaly begins, a regime switch from a stationary to a non-stationary state should be detected, so that $\pi_t^{\mathtt{ns}}$ should go from a value near zero to a value near one. For the agent, the term *anomaly detection* refers to taking an action from a set $\mathcal{A} = \{a^0 : \textit{trigger an alarm}, a^1 : \textit{not trigger an alarm}\}$, depending on the presence of an anomaly; the agent's action should be triggering an alarm at a state $\mathbf{s}_{t > t_{\mathrm{s}}}$ such as the one shown by the red dotted circle, and it should be as close as
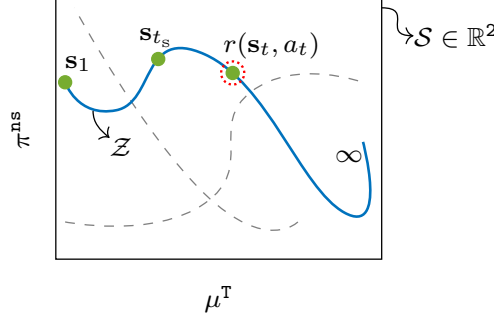
Figure 4.1 Illustration of the episodes $\mathcal{Z}$ and environment's state vector $\mathbf{s}_t = [\mu_t^\mathsf{T} \; \pi_t^{\mathtt{ns}}]^\mathsf{T} \in \mathcal{S}$ for which the agent triggers an alarm at time t during the non-stationary regime. The curves inside the environment represent different realizations. One of them (solid line) is highlighted to explain the nomenclature; $\mathbf{s}_{t_\mathrm{s}}$ is associated with the start time of an anomaly. The red dashed circle indicates the time for which the agent triggers the alarm.

possible to the time t_s. On the other hand, for the times $t < t_\mathrm{s}$, the agent's action should be not triggering an alarm. Here, once the agent selects the action of triggering an alarm at any time, the episode is terminated and restarted with a new one as shown by dashed lines in Figure 4.1. The reward $r(\mathbf{s}_t, a_t) \in \mathbb{R}$ is a scalar function of the environment's state $\mathbf{s}_t$ and of the agent's action a_t indicating whether the agent takes desirable actions during the agent-environment interactions [67]. In Section 4.1.2, we present the reward function and its values. A *policy* $\pi : \mathbf{s} \to a$ is a map from a state to an action to be taken by an agent. For a given policy π, the value of taking the action a in the environment's state $\mathbf{s}$ is denoted by the action-value function $q_\pi(\mathbf{s}, a)$ as shown by Equation 2.21 in Section 2.3.1. In this setup, the anomaly detection can be regarded as a sequential decision making problem, where the objective is to maximize the long-term expected return quantified by the action-value function. Therefore, the anomaly detection problem is governed by the execution of the optimal policy π^* presented in Equation 2.22 following

$$
\begin{aligned}
\pi^*(\mathbf{s}) &= \arg\max_{a_t} \; q_{\pi^*}(\mathbf{s}_t, a_t), \\
&= \arg\max_{a_t} \; \mathbb{E}_{\pi^*}\left[\sum_{t=1}^{\infty} \gamma^{t-1} r_t(\boldsymbol{S}_t, a_t) \right].
\end{aligned} \tag{4.1}
$$

RL-based anomaly detection has two advantages; First, the policy depends on the environment's state $\mathbf{s}$, which consists in the information from both the local trend hidden state variable and the probability of the non-stationary regime. This addresses the issues regarding the decision making being solely based on the probability of the non-stationary regime outlined in Section 2.2.5. Second, in Equation 4.1, evaluating the action-values depends on

the current and future reward values. This implies that the decision making considers the long-term effect of taking actions. Therefore, maximizing the accumulated discounted rewards results in maximizing the anomaly detectability, while minimizing the number of false and missed alarms. Note that in the aforementioned setup, the agent does not control the state variables. Rather, it follows the trajectories, such as the one shown in Figure 4.1, which are obtained from the BDLM analysis of the time series.

We use the Q-learning method in order to train an agent in a discretized environment's state-space $\hat{\mathcal{S}}$ such that $d_{\texttt{M}\times\texttt{N}} : \mathcal{S} \to \hat{\mathcal{S}}$ is a discretization map from a continuous environment's state-space $\mathcal{S}$ into a discretized one $\hat{\mathcal{S}}$ with a grid size $\texttt{M} \times \texttt{N}$. Figure 4.2 shows the same environment's state-space depicted in Figure 4.1 which has been discretized with a grid size of 5×4.
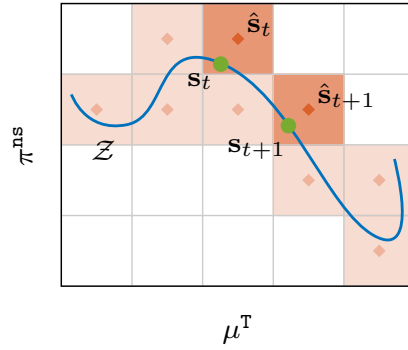
Figure 4.2 Mapping the continuous environment's state $\mathbf{s}_t = [\mu_t^{\texttt{T}} \ \pi_t^{\texttt{ns}}]^{\texttt{T}} \in \mathcal{S}$ into discretized one $\hat{\mathbf{s}}_t \in \hat{\mathcal{S}}$ for two successive times t and $t+1$ during an episode $\mathcal{Z}$. The shaded areas and diamonds marks indicate the discretized regions and their centers visited by the agent. The darker shaded areas correspond to the regions visited by the agent for a consecutive time steps t and $t+1$.

Two successive continuous environment's states $\mathbf{s}_t$ and $\mathbf{s}_{t+1}$ are shown by solid circles; the corresponding discretized states are respectively $\hat{\mathbf{s}}_t = d(\mathbf{s}_t)$ and $\hat{\mathbf{s}}_{t+1} = d(\mathbf{s}_{t+1})$, which are shown by darker shaded areas. In general, the grid size depends on the problem at hand. Using a coarse grid may result in poor policies for the discretized environment's states. Therefore, the grid needs to be fine enough to ensure the convergence of the Q-values [85]. On the other hand, increasing the grid size increases the number of states and the subsequent memory allocations during the learning phase; This results in a slower learning as well as the necessity for using a larger amount of data to train an agent. We set the grid size based on the pilot study we carried out on the Champlain Bridge, Canada [56]. Here, we use a fine grid size for lower expected values of the local trend and lower values of the probability

of the non-stationary regime; This allows the agent to be trained for small changes in the environment associated with these low values. On the other hand, for higher values of the mean trend an the probability of the non-stationary regime, we use a coarse grid.

In a discretized RL setup, the task is to find the optimal policy by interacting with the environment through several episodes as shown in Figure 4.1. Each episode includes $\mathtt{T}$ time steps and at the end-time of each episode, i.e., $t = t_{\mathtt{T}}$, the agent cannot select further action to determine the next state. Consequently, at time $t = t_{\mathtt{T}-1}$, the agent-environment interaction is terminated and restarted with another episode such as the ones depicted in Figure 4.1 by dashed lines. The optimal policy is attained when the Q-values converge to stationary values. Here, the Q-values are considered as having converged when the expected Q-values over all visited states for $\mathtt{K}$ consecutive episodes satisfy the *convergence error ratio*

$$\delta_{i:i+\mathtt{K}} = \frac{\left| \mathbb{E}_{\hat{\mathbf{s}}}[Q_{i+k}(\hat{\mathbf{s}})] - \mathbb{E}_{\hat{\mathbf{s}}}[Q_i(\hat{\mathbf{s}})] \right|}{\left| \mathbb{E}_{\hat{\mathbf{s}}}\left[Q_i(\hat{\mathbf{s}})\right] \right|} \leq \delta_0, \ \forall k = \{1, 2, \cdots, \mathtt{K}\}, \tag{4.2}$$

where δ_0 is a user-defined convergence tolerance and $Q_i(\hat{\mathbf{s}})$ is the Q-value for a state $\hat{\mathbf{s}}$ at time $\mathtt{T} - 1$ of the ith episode. The subscript $\hat{\mathbf{s}}$ indicates that the expected value is computed from all the visited states. Note that, the choice of the grid size was studied via a grid search in order to ensure the convergence of the learning process following Equation 4.2. In this thesis, the expected Q-values over all visited states is computed for ten consecutive episodes, i.e., $\mathtt{K} = 10$, and the user-defined convergence tolerance δ_0 is set to 1×10^{-6}.

4.1.1 Environment's State-Space

For structural health monitoring applications, agent-environment interactions should mimic the underlying structural responses in the presence of various anomalies. Training an agent to detect anomalies solely based on real data cannot address this challenge because anomalies are rare events in the context of infrastructure monitoring. This limits training an agent to learn from various anomalies. Therefore, we use the BDLM models built in order to simulate new time series. Afterwards, we employ the newly generated time series along with a stochastic *anomaly function* in order to build an environment's state-space which with an agent interacts.

Stochastic Anomaly Function

We define an anomaly as a switch from a stationary to a non-stationary regime. In order to simulate such a transition, we define an anomaly function representing a change in the

structure's degradation using a continuous logistic function

$$f_{\mathrm{a}}(t; \mathcal{J}) = \frac{h_{\mathrm{a}}}{|h_{\mathrm{a}}|} \cdot \frac{|h_{\mathrm{a}}| + 2\epsilon_{\mathrm{a}}}{1 + e^{-k(t-t_{\mathrm{c}})}}, \tag{4.3}$$

in which the steepness k is governed by

$$k = \frac{2}{w_{\mathrm{a}}} \ln\left(\frac{|h_{\mathrm{a}}|}{\epsilon_{\mathrm{a}}} + 1\right). \tag{4.4}$$

A logistic function is differentiable over its entire domain and have exact closed-form derivatives, which determines the kinematics of the anomaly function. In Equation 4.3, the logistic function parameters can be generated from $\mathcal{J} = \{h_{\mathrm{a}} : H_a,\ w_{\mathrm{a}} : W_a,\ t_{\mathrm{c}} : T_c\}$, a set of random variables describing the stochastic characteristics of a non-stationary regime. It consists in the anomaly's magnitude h_{a}, its duration w_{a}, and the center time t_{c} of the anomaly function. The user-defined *anomaly tolerance* ϵ_{a} is the minimum magnitude corresponding to the time for which a non-stationary regime starts, i.e., $f_{\mathrm{a}}(t_{\mathrm{s}}) = \epsilon_{\mathrm{a}}$.

Figure 4.3 illustrates three different anomaly realizations with the same magnitude $h_{\mathrm{a}} > 0$ and different centers and durations such that $w_{\mathrm{a}}^1 < w_{\mathrm{a}}^2 < w_{\mathrm{a}}^3$. Hatched regions on the horizontal axis indicate the out-of-domain values which are discarded. For the anomaly function shown by a solid line, the non-stationary duration w_{a} is centered at the time t_{c} and is associated with the start time $t_{\mathrm{s}} = t_{\mathrm{c}} - w_{\mathrm{a}}/2$ and the end time $t_{\mathrm{e}} = t_{\mathrm{c}} + w_{\mathrm{a}}/2$. Anomaly function values corresponding to the start and end time of the non-stationary duration are ϵ_{a} and
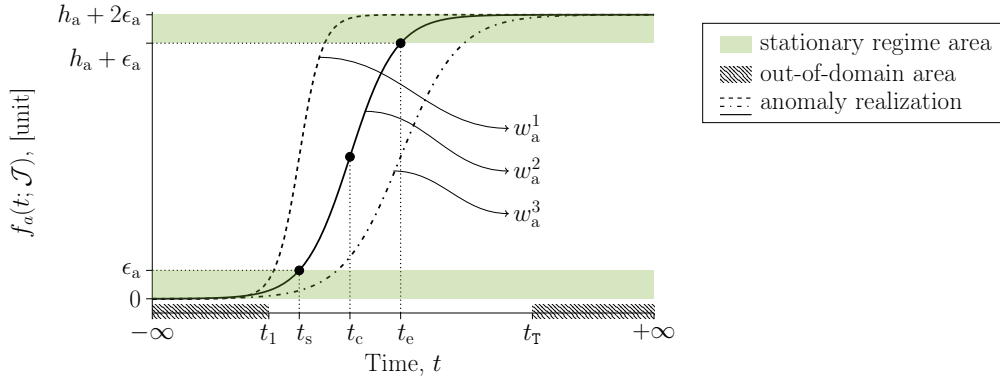
Figure 4.3 Illustration of an anomaly function $f_{\mathrm{a}}(t; \mathcal{J})$ with $h_{\mathrm{a}} > 0$. The curves represent three anomaly realizations with the same height $h_{\mathrm{a}}^1 = h_{\mathrm{a}}^2 = h_{\mathrm{a}}^3 = h_{\mathrm{a}}$ and anomaly tolerance $\epsilon_{\mathrm{a}}^1 = \epsilon_{\mathrm{a}}^2 = \epsilon_{\mathrm{a}}^3 = \epsilon_{\mathrm{a}}$, and different durations such that $w_{\mathrm{a}}^1 < w_{\mathrm{a}}^2 < w_{\mathrm{a}}^3$. One of them (solid line) is highlighted to explain the nomenclature. The non-stationary regime is centered at time t_{c} for which $t_{\mathrm{e,s}} = t_{\mathrm{c}} \pm w_{\mathrm{a}}/2$.

$h_{\mathrm{a}} + \epsilon_{\mathrm{a}}$, respectively. The green shaded areas indicate the anomaly tolerance for which the regime is considered to be stationary. For the cases where $|h_{\mathrm{a}}| < \epsilon_{\mathrm{a}}$ or $h_{\mathrm{a}} = 0$, we assume that there is no anomaly, i.e., $f_{\mathrm{a}}(t) = 0$, $\forall t$. In Figure 4.3, the anomaly function magnitude $|f_{\mathrm{a}}(t)| \in (0, |h_{\mathrm{a}}| + 2\epsilon_{\mathrm{a}})$ approaches 0 as t tends to $-\infty$. In addition, an anomaly starts some time steps after the time t_1 denoting the first timestamp of the an episode. Therefore, at the time t_1, the anomaly function introduces an error of magnitude $|f_{\mathrm{a}}(t_1)|$, which must satisfy $|f_{\mathrm{a}}(t_1)| < \varepsilon$. Here, ε is a user-defined tolerance. In order to start episodes with a stationary regime, we accept anomaly realizations for which $t_{\mathrm{s}} \geq t_1 + w_0$, where $w_0 > 0$ is the number of initial time steps without any anomaly.

Simulated Environment's State Space

The process of simulating an environment's state-space involves generating synthetic abnormal structural responses and then estimating hidden-state variables and regime probabilities; Generating abnormal structural responses is carried out by simulating synthetic time series using the BDLM matrices associated with a normal regime $\mathcal{M}^{\mathrm{s,obs}}$, and the prediction step of the Kalman filter as explained in Section 2.2.2. Afterwards, the values $f_{\mathrm{a}}(t; \mathcal{J})$ for a realization of the anomaly function are added to the synthetic time-series. This process is repeated in order to build synthetic times series containing a wide range of anomalies. The simulation of synthetic structural responses $\mathbf{y}_t^{\mathrm{sim}}$ is summarized by

$$\mathbf{y}_t^{\mathrm{sim}} = \mathrm{Sim}(\boldsymbol{\mu}_{t|t}^{\mathrm{obs}}, \boldsymbol{\Sigma}_{t|t}^{\mathrm{obs}}, \mathcal{M}^{\mathrm{s,obs}}, \mathcal{P}^{*,\mathrm{obs}}, f_{\mathrm{a}}(t; \mathcal{J})), \tag{4.5}$$

where, the superscripts $^{\mathrm{obs}}$ and $^{\mathrm{sim}}$ refer to the quantities that are obtained from the observed and simulated structural responses, respectively. Once abnormal synthetic structural responses are obtained, we employ the switching Kalman filter, presented in Section 2.2.2, in order to estimate its hidden-state variables and regime probabilities.

Figure 4.4a shows an example of ten synthetic responses containing anomalies. In each figure, the red line is the actual observations $\mathbf{y}^{\mathrm{obs}}$. The black lines are the simulated responses $\mathbf{y}^{\mathrm{sim}}$ for 10 years for which $t_1 = 2019\text{-}08$ and $t_{\mathrm{T}} = 2029\text{-}08$. The vertical dashed line indicates the initial number of time steps in days without anomaly, i.e., $w_0 = 90$. For a better illustration, Figure 4.4b shows the synthetic time series and observations corresponding to Figure 4.4a for a shorter time span that is between 2019-08 and 2022-02. The synthetic anomalies are generated by sampling the anomaly function parameters from the corresponding uniform distributions. The types of distributions and the corresponding parameters for generating anomalies depend on the requirements of the anomaly detection with respect to the prior knowledge about an anomaly's magnitude and duration.
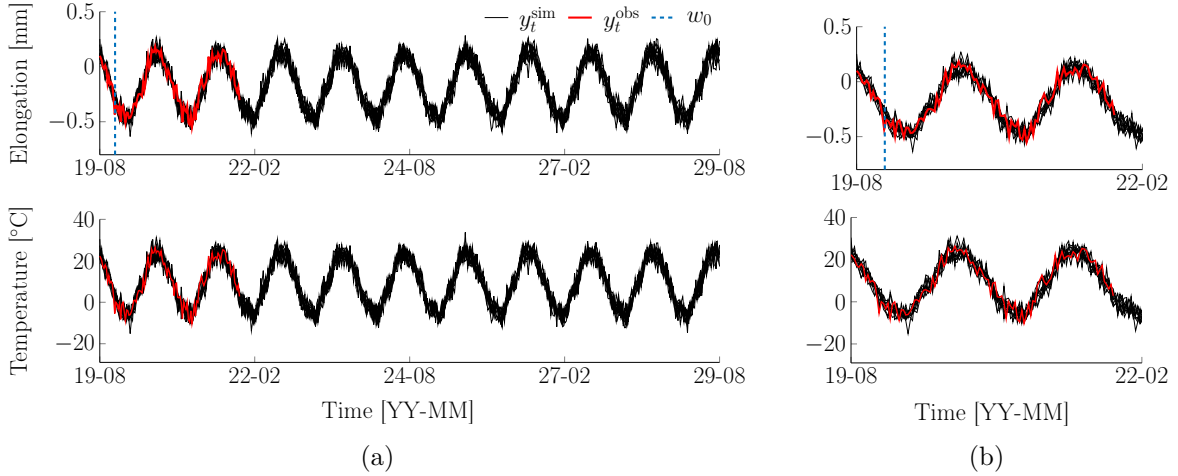
Figure 4.4 Comparison between 10 synthetic abnormal structural responses $\mathbf{y}^{\text{sim}}$ and observations $\mathbf{y}^{\text{obs}}$ for elongation and temperature measurements. The anomaly's height, duration, and the time corresponding to the center of the non-stationary regime are sampled from the respective uniform distributions $H_{\text{a}} \sim \mathcal{U}(h_{\text{a}}; -0.03, 0.03)\,\text{mm}$, $W_{\text{a}} \sim \mathcal{U}(w_{\text{a}}; 0, 5)\,\text{year}$, and $T_{\text{c}} \sim \mathcal{U}(t_{\text{c}}; 2019\text{-}08, 2029\text{-}08)$. The number of time steps without synthetic anomaly $w_0 = 90$ days: (a) ten years synthetic time series simulations, and (b) the same synthetic time series and observations for a shorter time between 2019-08 and 2022-02.

Figure 4.5a illustrates the estimated local trend expected values and the probability of the non-stationary regime for the ten synthetic structural responses shown in Figure 4.4. The vertical yellow dash-dotted lines indicate the anomalies' center time t_{c}. The shaded areas correspond to the duration w_{a} of the anomalies. Figure 4.5b shows the simulated environment's state-space in which we presented 10 episodes corresponding to the estimated values shown in Figure 4.5a. In Figure 4.5, two episodes are emphasized by the green and red colors. They respectively show high and low probabilities of detecting anomalies. Section 2.2.5 presented an example of how relying only on the probability of the non-stationary regime to detect anomalies is prone to false or missed alarms. The RL-based anomaly detection method copes with this issue by incorporating the trend hidden-state variable along with the probability of the non-stationary regime. Moreover, we do not have a precise prior knowledge about the occurrence of anomalies in the sense that we do not know when, with what magnitude, and with what duration anomalies will happen. The anomaly detection method proposed thus uses a wide range of anomalies during the training phase such as the two examples shown in the green and red colors in Figure 4.5.

Figure 4.6a shows the continuous states $\mathbf{s}$ (blue dots) and their corresponding discretized states $\hat{\mathbf{s}}_t$ (shaded regions) associated with 100 episodes obtained from the simulation process with the same configuration that was used to generate the states in Figure 4.5.

(a) (b)
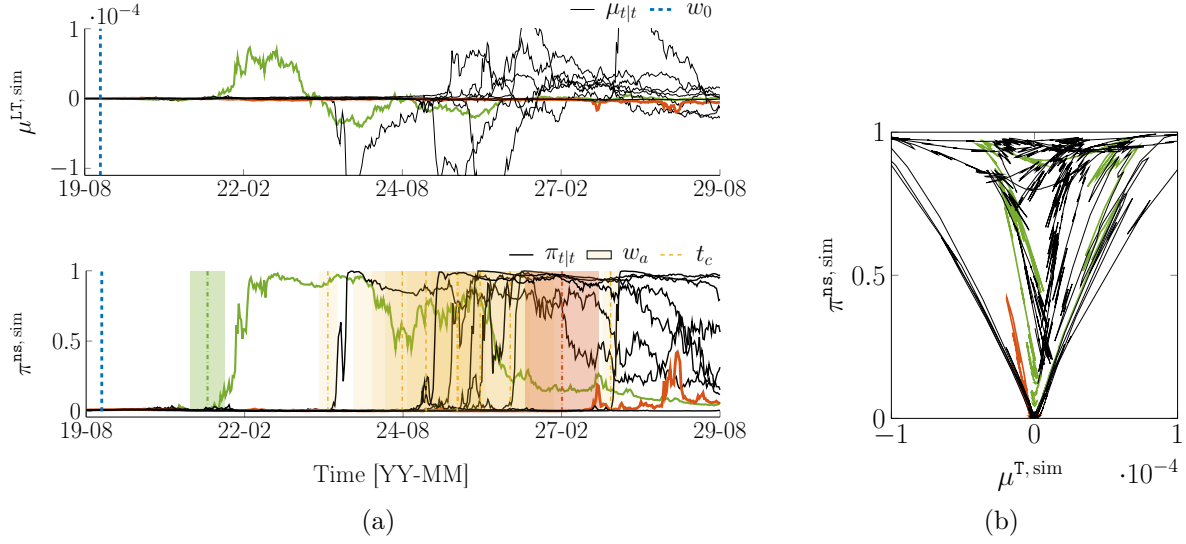
Figure 4.5 Simulation of 10 episodes: (a) illustration of the abnormal episodes. The vertical dash-dotted lines indicate the anomalies' center time t_c. The shaded areas correspond to the duration w_a of the anomalies, and (b) illustration of the environment's state-space $\mathcal{S}$.

(a) (b) (c)
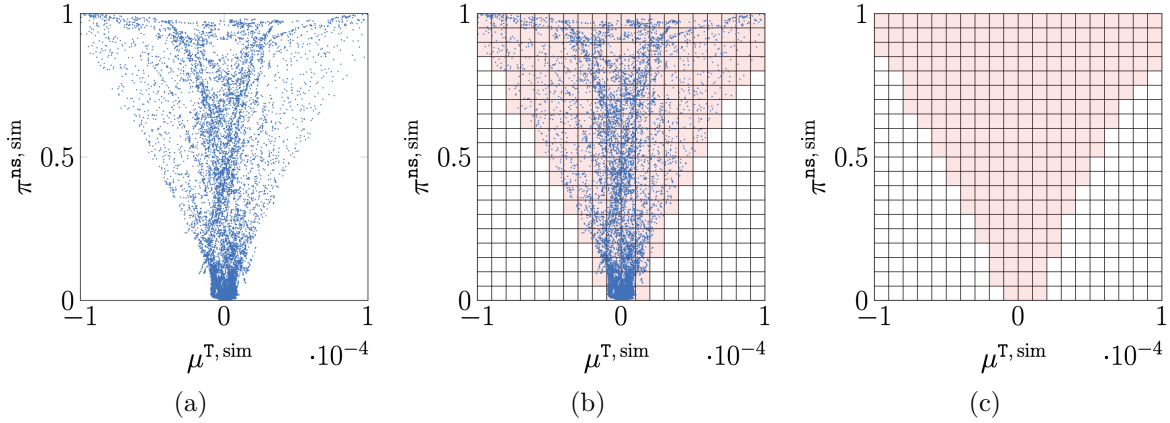
Figure 4.6 Example of a discretized environment using 20×20 grid along with the data points (blue dots) corresponding to 100 episodes mapped on the environment's states: (a) the continuous states $\mathbf{s}$ obtained from 100 episodes with the same configuration that was used to generate the states in Figure 4.5, (b) the 20×20 grid along with the mapped data into the 400 induced discretized states states $\hat{\mathbf{s}}$. The shaded regions represent the visited states $\hat{\mathbf{s}}$, while the while ones are associated with non-visited states $\neg\hat{\mathbf{s}}$, and (c) the final discretized environment for training an agent. The shaded regions are the discretized states for which an agent interacts with.

4.1.2 Reward Function

The reward function represents a feedback from the environment determining whether the agent should be encouraged or discouraged for the action it takes [86]. The reward function $r(\mathbf{s}_t, a_t)$ for an episode $\mathcal{Z}$ takes the form of a *confusion matrix* as shown in Table 4.1. Reward values are based on the comparison between the optimal action and actions taken by an agent at the time t. Note that because each episode is issued from an anomaly simulation process, we know whether an anomaly is present or not. In Table 4.1, the *true positive* reward r^{TP} and *true negative* reward r^{TN} correspond to the environment's states where the agent correctly selects the optimal action of triggering an alarm in the presence of an anomaly or doing nothing otherwise. On the other hand, the *false positive* (false alarm) reward r^{FP} and *false negative* (delayed or missed alarm) reward r^{FN} are associated with an incorrect action taken by an agent.

Table 4.1 Reward function $r(\mathbf{s}_t, a_t)$ presented in the form of a confusion matrix. The action $a \in \mathcal{A} = \{a^0 : \ trigger\ an\ alarm, a^1 : \ not\ trigger\ an\ alarm\}$.

Agent action	Optimal action	
	a^0	a^1
a^0	r^{TP}	r^{FP}
a^1	r^{FN}	r^{TN}

In general, true positive and negative reward values are greater than the false positive and negative ones such that $r^{\mathrm{TN}}, r^{\mathrm{TP}} > 0$ and $r^{\mathrm{FP}}, r^{\mathrm{FN}} < 0$. The *delayed* reward $r^{\mathrm{FN}} = -\eta \hat{f}_{\mathrm{a}}(t; \mathcal{J})$, $t > t_{\mathrm{s}}$, depends on the normalized anomaly function $\hat{f}_{\mathrm{a}}(t; \mathcal{J}) \in (0, 1)$ such that the agent is discouraged incrementally as it delays triggering a rightful alarm; $\eta \geq 0$ is the *false negative factor* enabling the user to control the reward value.

4.1.3 Anomaly Detection Framework

The methods presented in previous sections are assembled into an anomaly detection framework applicable for full-scale infrastructure monitoring. The framework is illustrated in Figure 4.7 which encompasses three stages identified by the dashed boxes, namely Stage I, Stage II, and Stage III. They are respectively associated with the empirical model estimation, systematic abnormal time series generation, and employing the Q-learning method. The goal of the first two stages is to built an environment with which an agent interacts. To this end, the anomaly detection framework employs the empirical model of the structure to simulate
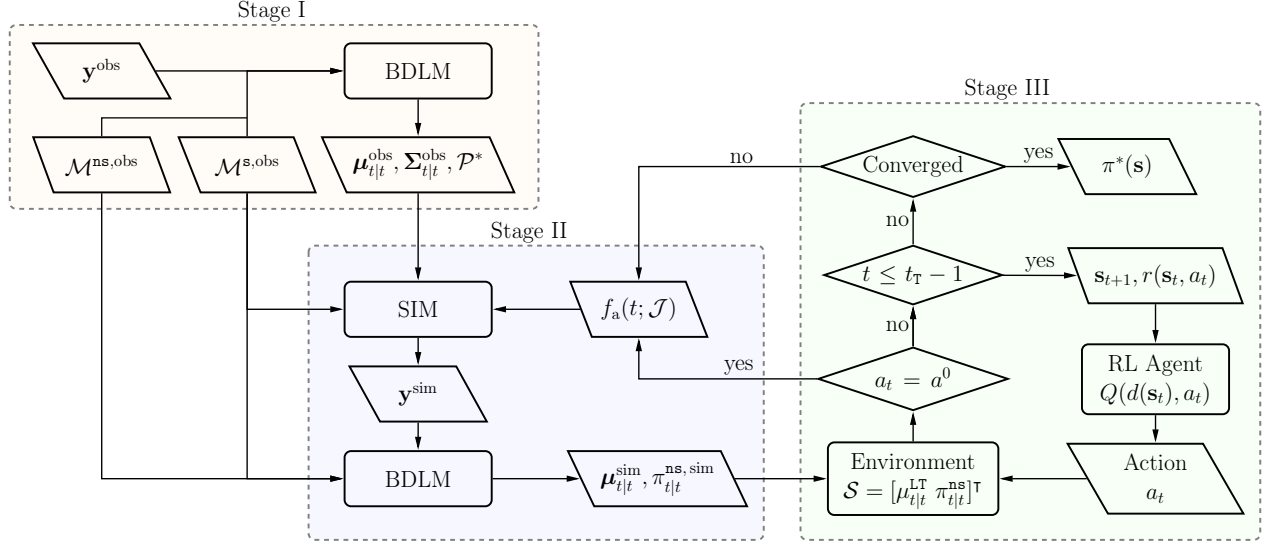
Figure 4.7 Illustration of the anomaly detection framework in the state-space model, which consists in three stages corresponding to the empirical model estimation (Stage I), systematic abnormal time series generation (Stage II), and employing the Q-learning method (Stage III).

episodes consisting in stationary and non-stationary regimes as described in Section 4.1.1; Stage I builds an empirical model from the structural responses $\mathbf{y}^{\text{obs}}$ and the user-defined generic components associated with the stationary and non-stationary regimes. The generic components involve model matrices $\mathcal{M}^{\text{s,obs}}$ and $\mathcal{M}^{\text{ns,obs}}$. In this stage, the BDLM uses the structural responses and model matrices to estimate the hidden state variables expected values $\boldsymbol{\mu}_{t|t}^{\text{obs}}$ and covariance matrix $\boldsymbol{\Sigma}_{t|t}^{\text{obs}}$, as well as the model parameters $\mathcal{P}^*$. In Stage II, the goal is to simulate new episodes from simulated structural responses $\mathbf{y}^{\text{sim}}$ as described in Section 4.1.1. To this end, structural responses are simulated from the stationary model matrix $\mathcal{M}^{\text{s,obs}}$, the stochastic anomaly function $f_{\text{a}}(t; \mathcal{J})$, and estimated hidden state variables expected values and covariance matrix obtained from the Stage I. Afterwards, BDLM uses both the model matrices and simulated structural responses to establish a new environment's state-space. The goal of the last stage is to train the Q-learning agent over multiple episodes in order to learn the alarm triggering policy as described in Section 2.3.1. In Stage III, two termination criteria are defined to restart the agent-environment interaction with a new episode by simulating new structural responses from Stage II. The first criterion, $t < t_{\text{T}} - 1$, is related to the end time of each episode, and the second one is related to the action taken by the agent such that if it triggers an alarm, the episode is terminated.

4.2 Probabilistic Quantification of Anomaly Detectability

This section presents the methods developed for quantifying the anomaly detectability with respect to two metrics, namely the annual false positive ratio and the probability of true positive detections.

4.2.1 Categorization of an Agent's Decisions

In order to quantify an agent's performance at detecting anomalies, we need to categorize the outcome of its decisions. Table 4.2 summarizes the four possible situations that an agent can be in, while following its policy $\pi(\mathbf{s})$. If an action is to trigger an alarm at any time t after the anomaly's start time t_s, the decision leads to a true positive (TP) detection, while if an agent triggers an alarm at any time step before t_s, it is a false positive (FP) one. In addition, if an action is to not trigger an alarm for any time step after t_s, it leads to a false negative (FN) detection, while for any time steps before t_s it is a true negative (TN) one. In order to draw probabilistic conclusions about the detectability of anomalies, we generate independent synthetic datasets with labeled anomalies and test an agent against them. As a result, we are able to quantify the anomaly detectability capacity of an agent with respect to two quantities: (i) the annual false positive ratio, and (ii) the probability of true positive detections.

Table 4.2 Agent detection terminology, given its policy $\pi(\mathbf{s})$ with the two possible actions a^0: trigger an alarm, and a^1: not trigger an alarm, and the time to detection with respect to the anomaly's start time t_s.

Terminology	Agent's action
True positive (TP)	$a^0,\ t \geq t_\mathrm{s}$
False negative (FN)	$a^1,\ t \geq t_\mathrm{s}$
True negative (TN)	$a^1,\ t < t_\mathrm{s}$
False positive (FP)	$a^0,\ t < t_\mathrm{s}$

4.2.2 Annual False Positive Detection Rate

The estimation of the annual false positive ratio ζ_FP involves testing the agent against an environment $\mathcal{S}^{\mathrm{test,FP}}$ for which all the episodes are free from anomaly, and then calculating the fraction of false positive detections from the *total number of false positive detections*. To this end, two quantities are evaluated in the test environment: (i) total number of time steps *#time steps* obtained from the total number of episodes in the environment, and (ii) the

average of the number of time steps per year *#time steps/year* for each episode. Therefore, the annual false positive ratio ζ_{FP} is expressed in events per year and is estimated by

$$
\begin{aligned}
\zeta_{\mathrm{FP}} &= \frac{\#\text{false positive detections} \times \#\text{time steps/year}}{\#\text{time steps}}, \\
&= \frac{\#\text{false positive detections}}{\text{year}}.
\end{aligned}
\tag{4.6}
$$

4.2.3   Probability of True Positive Detection

In order to evaluate the probability of a true positive detection, we only consider the actions for the time steps after the beginning of an anomaly, i.e, $t \geq t_{\mathrm{s}}$. This implies that only true positives and false negative detections are considered in order to build a probabilistic model for the true positive detections. The task of building such a probabilistic model corresponds to a classification problem with two classes $\{-1 \equiv \text{false negative}, +1 \equiv \text{true positive}\}$. The covariates for this classification problem are the absolute value of the anomaly's magnitude $|h_{\mathrm{a}}|$, its duration w_{a}, and the number of time steps $\mathtt{N}_{t>t_{\mathrm{s}}}$ that are required to detect an anomaly after its start. Note that the covariate $\mathtt{N}_{t>t_{\mathrm{s}}}$ determines the delay caused by an agent in detecting an anomaly, while the other two covariates are the anomaly's characteristics. For a test environment $\mathcal{S}^{\text{test, TP}}$, a set of D joint observations is defined by $\mathcal{D} = \{(\mathbf{x}_i, y_i), \forall i \in \{1 : \mathtt{D}\}\}$, where, $\mathbf{x}_i = [|h_{\mathrm{a}}| \ w_{\mathrm{a}} \ \mathtt{N}_{t>t_{\mathrm{s}}}]^{\mathsf{T}}$ are the covariates vector and $y_i \in \{-1, +1\}$ are the system responses. Note that whenever the policy is to trigger an alarm after the beginning of an anomaly, we consider that the policy for the remaining time steps of the episode to be triggering an alarm. This consideration is due to the termination criterion explained in Section 4.1.3, for which the agent-environment interaction is terminated whenever the agent triggers an alarm.

Figure 4.8 shows an example of an abnormal episode used in building a test data set $\mathcal{D}$. Figure 4.8a illustrates a realization of the anomaly function $f_{\mathrm{a}}(t; \mathcal{J})$. The anomaly's magnitude is $h_{\mathrm{a}} = 0.11$ and its duration is $w_{\mathrm{a}} = 1$ year. The unit of the anomaly magnitude is the same as the measurements'. Figure 4.8c depicts the resulting simulated time series using the anomaly realization along with the simulation procedure expressed in Equation 4.5. Figures 4.8b and 4.8d show the trend expected value μ^{T} and the probability of the non-stationary regime, respectively. As shown in Figure 4.8c, the agent triggers an alarm approximately 10 months $\equiv 0.83$ year after the anomaly begins. The point where the alarm is triggered is identified by a circle marker and a vertical dashed line.

Figure 4.9 shows an example of the results on a test environment containing 1200 episodes,

(a) Anomaly fucntion

(b) Expected value of the level

(c) Simulated time series

(d) Probability of the non-stationary regime
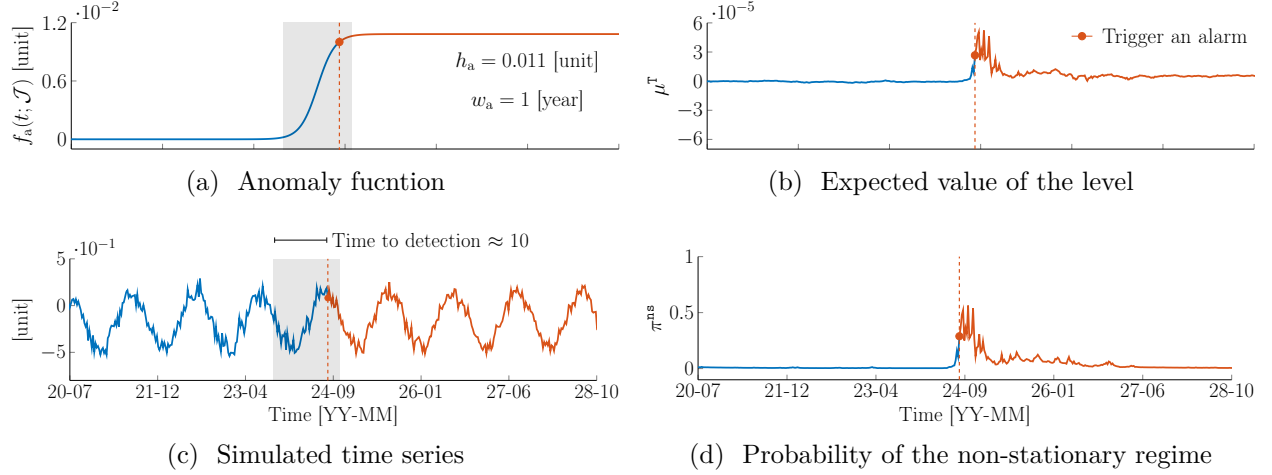
Figure 4.8 Example of an abnormal episode used to build a test data set $\mathcal{D}$. The realization of the anomaly function results in an anomaly magnitude of $h_a = 0.11$ and a duration of $w_a = 1$ year. The shaded area represents the anomaly duration for which the agent triggers an alarm approximately 10 months $\equiv 0.83$ year after the anomaly starts. The point where the alarm is triggered is identified by a circle marker and a vertical dashed line. The covariates vector is $\mathbf{x}_i = [0.011\ 1\ 0.83]^\intercal$.

where each consists in an anomaly having a different magnitude, duration, and time of occurrence. The three axes correspond to the covariates vector $\mathbf{x}$. Note that the unit of the
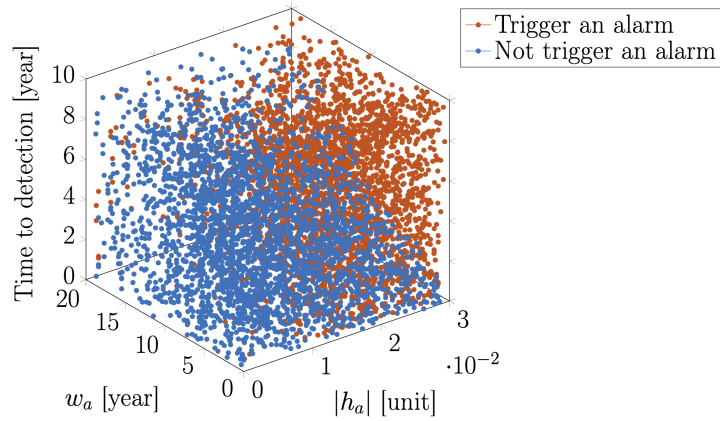
Figure 4.9 Example of the joint observations $\mathcal{D} = \{(\mathbf{x}_i, y_i), \forall i \in \{1 : \mathtt{D}\}\}$ from 1200 episodes, where the covariates vector $\mathbf{x}_i = [|h_a|\ w_a\ \mathtt{N}_{t>t_s}]^\intercal$, and the classes y_i are $-1 \equiv$ false negatives (blue dots) and $+1 \equiv$ true positives (red dots).

anomaly magnitude $|h_a|$ depends on the measurements' type that is mm for the extensometers and °d for the inclinometers. The red dots are associated with the true positive detections, where the action consists in rightfully triggering an alarm, i.e., $y = +1$. The blue dots

indicate the false negative detections, $y = -1$, for which the agent wrongfully avoids to trigger an alarm. Note that the episodes resulting in false positives are discarded because they are considered separately for evaluating the annual false positive ratio as explained in Section 4.2.2. We employ the *tractable approximate Gaussian inference* (TAGI) method [87] for Bayesian neural networks in order to build a classification model that will characterize the probability of a true positive detection as a function of the anomaly's magnitude, its duration, and the time to detection. TAGI is parameterized by the neural network's weights and biases. Once these parameters are estimated from the dataset $\mathcal{D}$, the probability of each class for a given covariates vector $\mathbf{x}$ can be evaluated. For the problem at hand, the probability of the two classes are: (i) the probability of a true positive detection $\Pr(Y = +1|\mathbf{x})$, conditional on the covariates vector $\mathbf{x}$, and (ii) the probability of a false negative detection $\Pr(Y = -1|\mathbf{x}) = 1 - \Pr(Y = +1|\mathbf{x})$.

Figure 4.10 illustrates the results obtained from the TAGI classification. Figure 4.10a presents an example of a model trained using TAGI with the observations obtained for 12000 episodes corresponding to the case presented in Figure 4.9. The colors correspond to the probability of true positive detections for any instance of the covariates vector $\mathbf{x}$. Figure 4.10b shows the isosurface extracted from Figure 4.10a for a 90% probability of true positive detections. Such an isosurface enables quantifying the anomaly detectability of an agent with respect

(a)   (b)
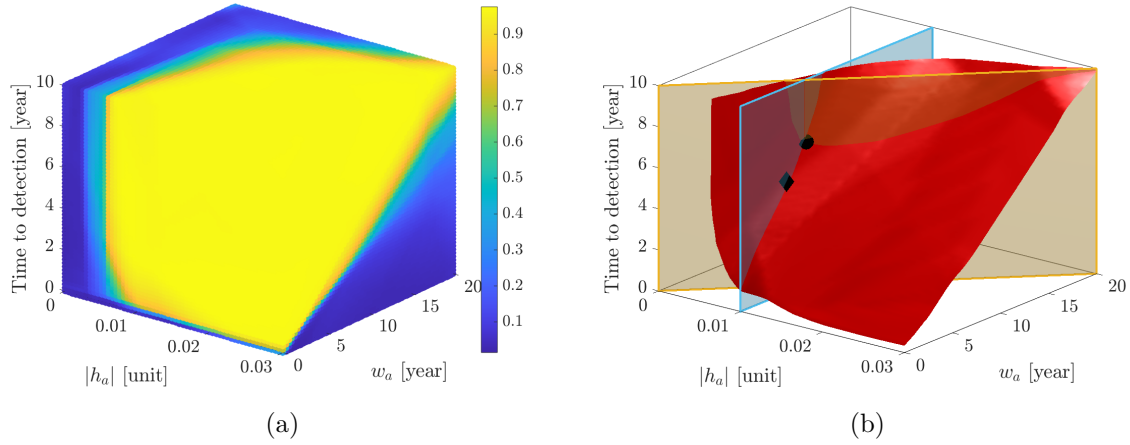
Figure 4.10 Example of a model trained using TAGI: (a) probability of true positive detections with respect to prediction covariates vector $\mathbf{x}$, and (b) the isosurface corresponding to 90% probability of true positive detections. The vertical planes indicate the minimum detectable anomaly magnitude $|h_a|_{\min} = 0.01$ (blue plane) and the minimum detectable anomaly growth rate per year $\psi_{\min} = 0.15|h_a|_{\min}$ (yellow plane). The black circle and diamond identifies two point located at the intersection of the isosurface and the vertical blue plane.

to two quantities identified by the two vertical planes: (i) the *minimum detectable anomaly magnitude* $|h_a|_{\min}$, which is shown by the blue color plane, and (ii) the *minimum detectable anomaly growth rate per year* $\psi_{\min} = \psi_0 \times |h_a|_{\min}$, which is shown by the yellow color plane. $\psi_{\min}$ is a function of the minimum detectable anomaly magnitude, where ψ_0 is a constant obtained from the slope of this plane with respect to the anomaly duration w_a. In Figure 4.10b, the minimum detectable anomaly magnitude $|h_a|_{\min}$ is 0.01. In addition, the minimum detectable anomaly growth rate is $\psi_{\min} = 0.15|h_a|_{\min}$ unit per year. For this case, the value of ψ_0 indicates that the anomaly with the minimum magnitude of 0.01 is detectable if its growth rate is 15% of the minimum detectable anomaly magnitude per year or higher. In other words, the anomaly with the magnitude of 0.01 is detectable with the probability of 90%, if it fully develops in less than $\mathbb{N}_\psi = 1/0.15 \approx 7$ years. For instance, the black circle in Figure 4.10 is located at the intersection of the two planes on the isosurface. It represents an anomaly with the minimum magnitude and the duration of 7 years. Such an anomaly is detectable within 7 years after it starts, with a probability of 90%. However, if this anomaly develops in more than 7 years, i.e., $w_a > 7$, the agent would not be able to detect it. Another interpretation of using the yellow plane is related to the point shown by a black diamond in Figure 4.10b. This point shows an anomaly with the minimum magnitude of 0.01 and the duration of roughly 5 years. As it is seen, this anomaly is detectable with a probability of 90% within 6 years after its starting time. However, the anomaly's magnitude for this point could also be associated with the anomaly shown by the black circle which is detected by the agent after 6 years. This discussion is also valid for other anomalies with the same magnitude but different durations. It is worth noting that the choice of the aforementioned plans is to provide simplified metrics in order to compare the anomaly detectability of each sensor using the reinforcement learning approach. Furthermore, in Chapter 5, these planes will be used as a basis to compare the performance of using reinforcement and imitation learning approaches. However, more complex functions could be used to carry out such comparisons depending on the engineering needs.

Such a probabilistic quantification of the anomaly detectability is crucial for the long term monitoring of infrastructures because anomalies with various magnitudes can develop over several years and an agent should detect them before they are fully developed. The next section presents the comprehensive results and discussions associated with the probability of true positive detections and the resulting isosurfaces for three instrumented spans of the Jacques Cartier Bridge.

4.3 Results

In this section, first we present the reinforcement learning agents used to detect anomalies on the three instrumented spans of the Jacques Cartier Bridge. The differences in these agents lie within their environments and reward values. Afterwards, we present the anomaly detectability of the agents with respect to the annual false positive ratio as well as the probability of true positive detections.

4.3.1 Jacques Cartier Bridge Reinforcement Learning Agents

This section presents the configurations used in order to train and test the reinforcement learning agents. The differences between the agents lie within the simulated environments that they interact with, and their reward values. The former depends on the sensor type, while the latter controls the annual false positive ratio. We target two annual false positive ratios $\zeta_{\mathrm{FP}} = 0.05$ and 0.1 per year that correspond to one false positive per 20 and 10 years, respectively. Note that the study of different agents allows illustrating how the reward values control the anomaly detectability of an agent [56].

Simulated Environments

We follow the procedure described in Section 4.1.1 in order to simulate environments. Rather than simulating an environment for each sensor, we simulate one environment for each set of sensors collecting the same measurements type across all three spans. This generalization of the environments across sensors is possible because the superstructures of the three spans and their subsequent structural responses are similar. As a result, there are four environments, and consequently four agents, which correspond to the lateral inclinometers, the longitudinal inclinometers 1, the longitudinal inclinometers 3, and the extensometers.

Constructing an environment involves defining the parameters associated with the simulation of new episodes and with the discretization of the environment. Table 4.3 shows the distributions and the parameters of the anomaly function for simulating new episodes for each sensor type. The duration of each episode is 10 years which starts in 2019 and ends in 2029. All probability distributions used for simulating the episodes have a uniform probability density with a lower a and an upper b bounds. The bounds for the anomaly magnitude depend on the type of the sensor, while they remain the same for other parameters associated with the anomaly function regardless of the sensor type; The duration for an anomaly is drawn from an uniform distribution with the upper bound $b = 2000$ days (≈ 5 years). The center of the anomaly is drawn from a uniform distribution for which the bounds are associated

with the beginning and end times of an episode. Here, the anomaly tolerance ϵ_a depends on the realization of an anomaly; its value is equivalent to 25% of the simulated anomaly's magnitude, i.e., $\epsilon_\mathrm{a} = 0.25 H_\mathrm{a}$.

Figure 4.11 illustrates 100 simulated episodes for each sensor type according to the parameters detailed in Table 4.3. Figure 4.11d shows the simulated episodes for the extensometers. As it is seen, the simulated episodes are coinciding with each other because the empirical models for this type of sensor result in similar parameter values as well as initial values for the hidden state variables. However this is not the case for the inclinometers; Figures 4.11a-

Table 4.3 Probability distributions and their parameters associated with the the anomaly function for simulating new episodes. The only difference in values is related to the anomaly magnitude that depends on the sensor type.

Sensor type	Parameter	Value		Unit
		a	b	
Inclinometer	$H_\mathrm{a} \sim \mathcal{U}(h_\mathrm{a}; a, b)$	-0.005	$+0.005$	°d
Extensometer		-0.030	$+0.030$	mm
All types	$W_\mathrm{a} \sim \mathcal{U}(w_\mathrm{a}; a, b)$	0	2000	day
	$T_\mathrm{c} \sim \mathcal{U}(t_\mathrm{c}; a, b)$	Aug 15, 2019	Aug 13, 2029	date
	w_0	constant:	90	day
	ϵ_a	constant:	$0.25 H_\mathrm{a}$	mm or °d
	ε	constant:	1×10^{-7}	mm or °d

(a)

(b)

(c)

(d)
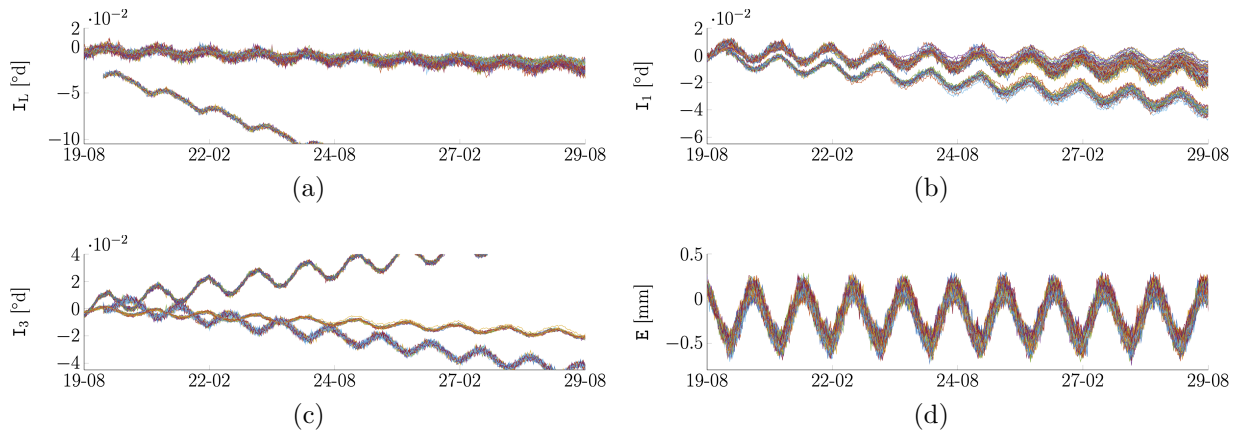
Figure 4.11 Illustration of 100 simulated episodes for each sensor type according to Table 4.3. The simulated episodes for each sensor type consist in the the episodes from the instrumented spans. The vertical dashed line indicates the number of days without anomaly, i.e., $w_0 = 90$ days. The sensor names are $\mathtt{I_L}$: Lateral inclinometer, $\mathtt{I_1}$: Longitudinal inclinometer 1, $\mathtt{I_3}$: Longitudinal inclinometer 3, and $\mathtt{E}$: Extensometer.

4.11c show that the simulated episodes for the inclinometers do not coincide with each other because either the estimated model parameter values or the initial values for the hidden state variables are different across the spans. Such differences are not an issue for training the agents because they are not trained on the time series data, rather they use the expected value of the trend hidden state variable and the probability of the non-stationary regime. In addition, the issue corresponding to the different initial trend expected values is resolved by removing the initial trend values $\mu_0^{\mathtt{T}}$ from each episode as shown in Equation 4.7.

Training and Test Datasets

Table 4.4 summarizes the dataset properties used in order to train and test each agent corresponding to a sensor type. For the training dataset $\mathcal{D}^{\mathrm{train}}$, we have simulated 400000 episodes associated with each sensor and then merged them into a single set of 1.2 million episodes for each sensor type. In addition, the simulated episodes for each sensor consist on average in 522 pairs of the trend expected value and the probability of the non-stationary regime, i.e, $(\mu_t^{\mathtt{T}}, \pi_t^{\mathtt{ns}})$, $t = \{1, 2, \ldots, \mathtt{T} = 522\}$. Therefore, the total number of continuous states exceeds 600 millions during the training. The last column of Table 4.4 indicates the ratio of abnormal episodes. For instance, 0.9 indicates that 90% of the total number of episodes during the training contain an anomaly simulated with the configuration presented in Table 4.3, while 10% of the total number of episodes do not contain any anomaly.

Table 4.4 Summary of the datasets used in order to train ($\mathcal{D}^{\mathrm{train}}$) and test ($\mathcal{D}^{\mathrm{test, FP}}$ and $\mathcal{D}^{\mathrm{test, TP}}$) each agent corresponding to its sensor type. The superscripts FP and TP are respectively associated with the evaluation of the annual false positive ratio and the probability of the true positive detections.

Data type	Notation	Episode			
		Total number	Duration, year	Anomaly	
				Duration, day	Ratio
Train	$\mathcal{D}^{\mathrm{train}}$	1.2×10^5	30	2000	0.9
Test	$\mathcal{D}^{\mathrm{test, FP}}$	3000	30	—	0.0
	$\mathcal{D}^{\mathrm{test, TP}}$	6000	60	8000	1.0

Simulating episodes for the test phase follows the same configurations as for the training phase shown in Table 4.3, except for the number and duration of episodes; In order to evaluate the annual false positive ratio, the test set $\mathcal{D}^{\mathrm{test, FP}}$ consists in a total number of 3000 episodes with a duration of 10 years for each measurement type. Note that the episodes for the false positive test set do not contain anomalies, i.e., $h_{\mathrm{a}} = 0$. Furthermore, the true positive test set

$\mathcal{D}^{\text{test, TP}}$ consists in 6000 abnormal episodes for each measurement type. The duration of each episode is 60 years, for which the anomaly duration in days follows a uniform distribution $W_{\mathrm{a}} \sim \mathcal{U}(w_{\mathrm{a}}; 0, 8000)$. Note that in Table 4.4, the superscripts FP and TP are respectively associated with the evaluation of the annual false positive ratio and the probability of the true positive detections

Discretized Environment

Table 4.5 shows the configurations for the two-dimensional environments with a grid size of $\mathtt{M} \times \mathtt{N}$, which are used to train the agents. The states of each environment are bounded according to the sensor type. All states have the same lower and upper bounds with respect to the probability of a non-stationary regime, i.e., $0 \leq \pi_t^{\mathtt{ns}} \leq 1$. The states' bounds differ with respect to the trend expected values, that is $|\mu_t^{\mathtt{T}}| \leq 5 \times 10^{-5}$ °d/day for the inclinometers, and $|\mu_t^{\mathtt{T}}| \leq 1 \times 10^{-4}$ mm/day for the extensometer. In addition all the environments are discretized using a grid of 160×60.

Table 4.5 Configurations of the discretized environments with a grid size of $\mathtt{M} \times \mathtt{N}$ that are used to train the agents corresponding to the for inclinometers and extensometers.

Sensor type	Lower bound		Upper bound		Grid size ($\mathtt{M} \times \mathtt{N}$)	
	$\mu_t^{\mathtt{T}}$ ($\times 10^{-4}$)	$\pi_t^{\mathtt{ns}}$	$\mu_t^{\mathtt{T}}$ ($\times 10^{-4}$)	$\pi_t^{\mathtt{ns}}$	$\mu_t^{\mathtt{T}}$ ($\mathtt{M}$)	$\pi_t^{\mathtt{ns}}$ ($\mathtt{N}$)
Inclinometer	−0.5	0.0	+0.5	+1.0	160	60
Extensometer	−1.0	0.0	+1.0	+1.0		

Agent-Environment Interaction

We use the Q-learning method described in Section 2.3.1 in order to train the agents. The hyperparameters for the exploration-exploitation method are $c_\alpha = c_\epsilon = 5$. In addition, we use the discount factor $\gamma = 0.97$ for all the agents, and the initial Q-value $Q_0 = 0$ for all the discretized states.

Furthermore, for the training of each agent, we take the absolute value of the trend expected values, i.e., $|\mu_t^{\mathtt{T}}|$ in order to reduce the computational cost. This is because mapping the simulated episodes on the environment results in a symmetrical set of visited states with respect to the dimension associated with the degradation's speed $\mu_t^{\mathtt{T}}$ and therefore, training only the positive half of the environment is sufficient. Moreover, in the case of inclinometers,

we remove the initial trend expected value μ_0^{T} from each episode following

$$\mathbf{s}_t = \left[(\mu_t^{\mathsf{T}} - \mu_0^{\mathsf{T}}) \ \pi_t^{\mathtt{ns}} \right]^{\mathsf{T}}. \tag{4.7}$$

This modification ensures that each episode begins from a state with a zero trend expected value because: (i) the initial trend for inclinometers are not equal, and (ii) without such a modification, we would not be able to take the advantage of the above-mentioned symmetrical property for the computational efficiency.

Figures 4.12 presents the probability contours for the visited states by the corresponding agents during the training phase for all the episodes in the training dataset. For instance, the red areas are associated with the states which are visited with a probability equal to 99%. As it is seen, many states are not visited, i.e., the white color regions, during the training because either an episode is terminated due to termination criterion described earlier, or because an episode does not visit these regions. For these non-visited states, the agent's action is considered to be to trigger an alarm, because they consist in high values of either the probability of the non-stationary regime or the expected value of the trend.
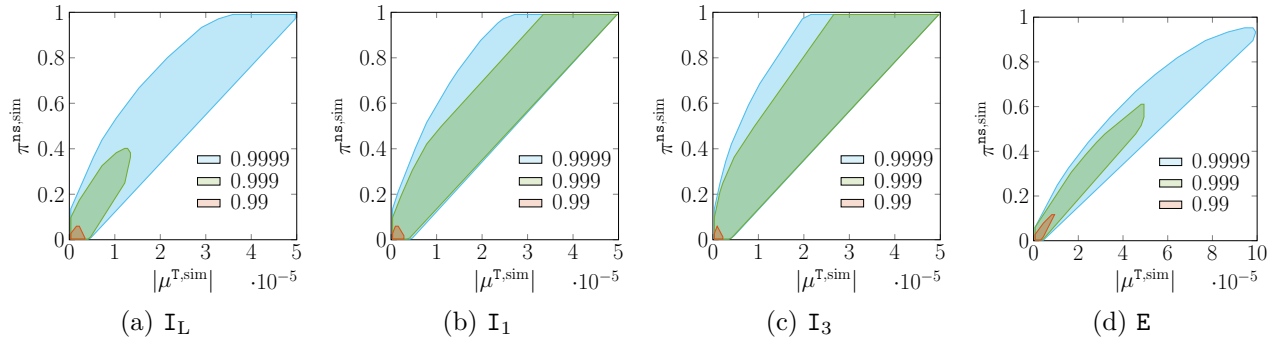
Figure 4.12 Examples of the probability of the visited states during the training phase. The three identified colors indicate the percentile of the agent-environment interactions. The non-visited states are shown by a white color. The sensor names are $\mathtt{I_L}$: Lateral inclinometer, $\mathtt{I_1}$: Longitudinal inclinometer 1, $\mathtt{I_3}$: Longitudinal inclinometer 3, and $\mathtt{E}$: Extensometer.

Reward Function Values

In order to achieve target annual false positive ratios of $\zeta_{\mathrm{FP}} = 0.05$ and 0.1 per year, we need to define specific reward values for each case. Table 4.6 shows the reward values for each sensor type where it can be noted that the false positive reward value r^{FP} associated with $\zeta_{\mathrm{FP}} = 0.05$ is less than the one for $\zeta_{\mathrm{FP}} = 0.1$. For instance, for the lateral inclinometer these values are -21 and -15 for the annual false positive ratio of 0.05 and 0.1, respectively. Such

Table 4.6 Reward values used to train the agents. Two reward sets are defined for each sensor type in order to obtain the annual false positive ratios of $\zeta_{\mathrm{FP}} = 0.05$ and 0.1 per year.

Sensor name	ζ_{FP}	Reward values			
		r^{TP}	r^{FP}	r^{TN}	η
$\mathtt{I_L}$: Lateral inclinometer	0.05	+4	−21	+1	+4
	0.10	+4	−15	+1	+4
$\mathtt{I_1}$: Longitudinal inclinometer 1	0.05	+4	−15	+1	+4
	0.10	+4	−9	+1	+4
$\mathtt{I_3}$: Longitudinal inclinometer 3	0.05	+4	−16	+1	+4
	0.10	+4	−10	+1	+4
$\mathtt{E}$: Extensometer	0.05	+4	−59	+1	+4
	0.10	+4	−26	+1	+4

a difference between the false positive rewards is because an agent must receive more penalty for false detections in the case of $\zeta_{\mathrm{FP}} = 0.05$ in order to keep the annual false positive ratio lower than 0.1. The values presented in Table 4.6 are obtained via a grid search approach where we kept all the reward values constant except the false positive reward in order to control the behavior of the agent with respect to false detections. Note that, it is possible to control the behavior of an agent via other reward values than only modifying r^{FP} [56]. The grid search results regarding the identification of appropriate reward values are provided in Appendix B.

4.3.2 Jacques Cartier Bridge Anomaly Detection

This section presents the optimal policies for the agents corresponding to the two cases resulting in the annual false positive ratios of $\zeta_{\mathrm{FP}} = 0.05$ and 0.1 per year. Hereafter, unless otherwise mentioned, we use a shorthand notation in order to simplify the name of the sensors following $\mathtt{I_L}$: Lateral inclinometer, $\mathtt{I_1}$: Longitudinal inclinometer 1, $\mathtt{I_3}$: Longitudinal inclinometer 3, and $\mathtt{E}$: Extensometer.

Figure 4.13 illustrates the resulting policies for the two annual false positive ratios. Due to the symmetrical property of the trend expected values, only the positive half of the environment is shown. The red color indicates that the action is a^0: Trigger an alarm, and the blue one indicates that the action is a^1: Not trigger an alarm. The white areas are associated with the non-visited states. In addition, we consider that the agents' decision is to trigger an alarm for the out-of-domain states with respect to the degradation speed, that is $|\mu_t^{\mathrm{T}}| > 5 \times 10^{-5}$ °d/day for the inclinometers and $|\mu_t^{\mathrm{T}}| > 1 \times 10^{-4}$ mm/day for the extensometers. This consideration is due to the large degradation speed of the out-of-domain states for which the probability

(a) $\mathtt{I_L}$, $\zeta_{\mathrm{FP}} = 0.05$/year   (b) $\mathtt{I_1}$, $\zeta_{\mathrm{FP}} = 0.05$/year   (c) $\mathtt{I_3}$, $\zeta_{\mathrm{FP}} = 0.05$/year   (d) $\mathtt{E}$, $\zeta_{\mathrm{FP}} = 0.05$/year

(e) $\mathtt{I_L}$, $\zeta_{\mathrm{FP}} = 0.10$/year   (f) $\mathtt{I_1}$, $\zeta_{\mathrm{FP}} = 0.10$/year   (g) $\mathtt{I_3}$, $\zeta_{\mathrm{FP}} = 0.10$/year   (h) $\mathtt{E}$, $\zeta_{\mathrm{FP}} = 0.10$/year
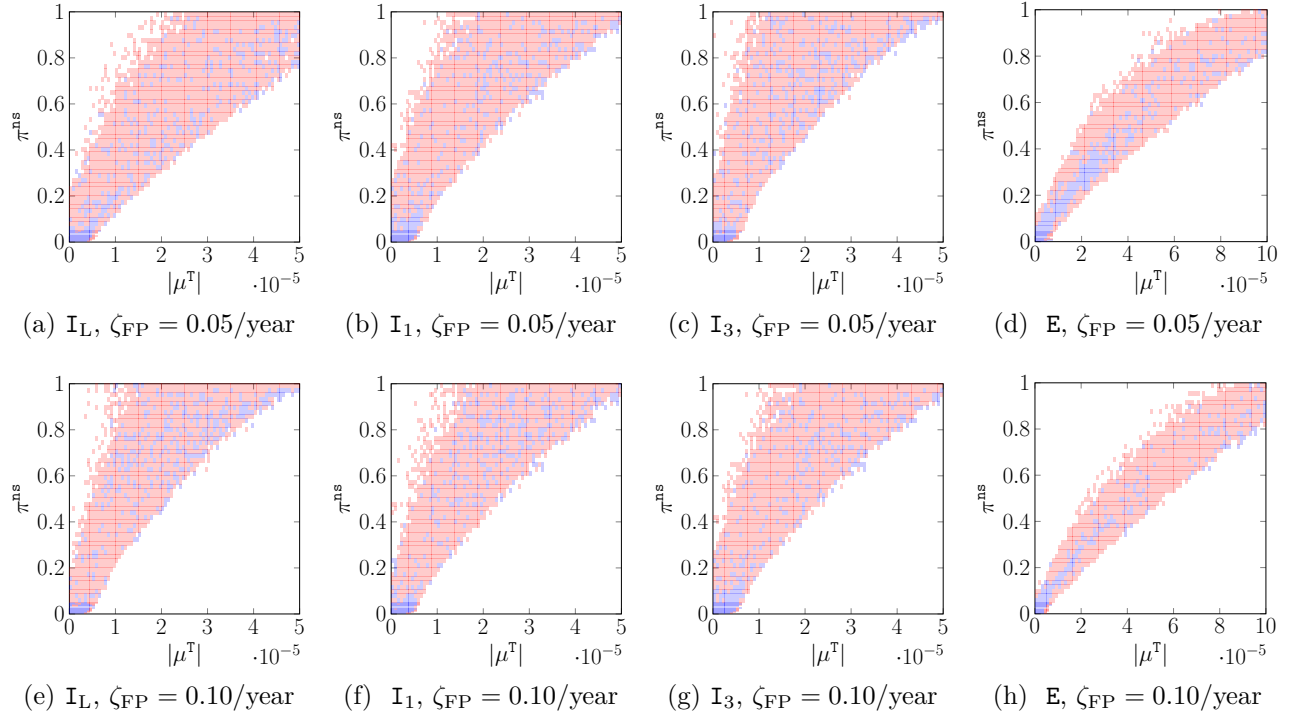
Figure 4.13 The policy maps for the inclinometers and the extensometers in two cases: (a-d) $\zeta_{\mathrm{FP}} = 0.05$/year, and (e-h) $\zeta_{\mathrm{FP}} = 0.10$/year. For a better illustration, the grid lines are not shown. The red color indicates that the action is a^0: Trigger an alarm, and the blue one indicates that the action is a^1: Not trigger an alarm. The white areas are associated with the non-visited states. Also, $\mathtt{I_L}$: Lateral inclinometer, $\mathtt{I_1}$: Longitudinal inclinometer 1, $\mathtt{I_3}$: Longitudinal inclinometer 3, and $\mathtt{E}$: Extensometer.

of the non-stationary regime is estimated by the BDLM to be close to one. This situation corresponds to the abnormal state of the structure for which the decision should be triggering an alarm, and therefore there is no need to train an agent for such a case. Rather, the training of the agents must be carried out for the situations where the probability of the non-stationary regime is less than one, yet we are not able to decide whether or not the situation represents an abnormal state of the structure. Figures 4.13a-4.13d correspond to the case for which the annual false positive ratio is $\zeta_{\mathrm{FP}} = 0.05$/year, while Figures 4.13e-4.13h are associated with the $\zeta_{\mathrm{FP}} = 0.10$/year. For a $(|\mu_t^{\mathrm{T}}|, \pi_t^{\mathrm{ns}})$ pair in a continuous space, each figure shows the policy of the agent in the discretized environment. Note that according to Table 4.4, the annual false positive ratio for each agent is evaluated on its corresponding independent test dataset $\mathcal{D}^{\mathrm{test, FP}}$ consisting in 3000 normal episodes each uses a 10 years duration. The agents' action is to do nothing for the region associated with low degradation speed and probability of the non-stationary regime, i.e., the blue region at the bottom left corners in Figures 4.13a-4.13h. However, this region is larger for the cases with the false positive ratio of $\zeta_{\mathrm{FP}} = 0.05$/year

compared with the ones with $\zeta_{\mathrm{FP}} = 0.1$/year. This difference is due to changes in the reward function values such that the higher is the penalty for r^{FP}, the larger the blue region; The penalties for the false positive decisions for the agents associated with $\zeta_{\mathrm{FP}} = 0.05$/year is higher than the ones for the agents related to $\zeta_{\mathrm{FP}} = 0.10$/year. As a result, the former agents have a tendency to trigger lesser alarms than the later ones. Other differences in the agents' behavior are sensor dependent. For instance, in the cases of the extensometer shown in Figures 4.13d and 4.13h, the blue region is narrower for the case associated with higher false positive ratio shown in Figure 4.13h, which indicates that the agent has the tendency to trigger more alarms for the same region compared with the agent shown in Figure 4.13d.

The above discussions are examples of how the reward function values control the behavior of an agent for a long-term decision making that is taking into account the annual false positive ratios. Choosing a lower penalty for wrongfully detecting anomalies increases the false positive ratio, and subsequently brings about more false alarms. The current work is limited to the three instrumented spans. Changing the false positive ratio from 0.05 to 0.1 for each span increases the number of false alarms for all three spans from approximately one every $20/3 \approx 7$ years to one every $10/3 \approx 3$ years. However, if all the 40 spans of the Jacques Cartier Bridge would be instrumented, above values would change from two false alarms per year to four false alarms per year. These calculations show the importance of being able to control the false alarm ratio within the anomaly detection methodology.

Before proceeding with quantifying the probability of true detections, we apply the policies obtained in Figure 4.13 on the real data collected from the sensors. Figures 4.14-4.16 illustrate the anomaly detection on the data collected from the inclinometers. In each figure, the points where alarms are triggered are shown with two markers; The yellow marker corresponds to the annual false positive ratio $\zeta_{\mathrm{FP}} = 0.05$/year, and the red one is associated with $\zeta_{\mathrm{FP}} = 0.10$/year. Note that the figures corresponding to the anomaly detection for the extensometers are not provided because the agents rightfully avoid triggering any alarm for these sensors.

Figure 4.14 shows the anomaly detection associated with the lateral inclinometers, i.e., $\mathtt{I_L}$, across all three instrumented spans. The results show that the agents corresponding to the lateral inclinometers identify the same timestamp where an alarm is triggered for both annual false positive ratios across the spans $5 - 6$ and $6 - 7$. However, for the span $4 - 5$ as shown in Figure 4.14a, the agent corresponding to the annual false positive ratio of $\zeta_{\mathrm{FP}} = 0.10$/year triggers an alarm three weeks after the agent associated with the annual false positive ratio of $\zeta_{\mathrm{FP}} = 0.05$/year. In general, we expect that the agent with higher false positive penalty has a tendency to delay the alarm triggering in order to achieve a lower annual false positive ratio.

(a) Span $4-5$

(b) Span $5-6$

(c) Span $6-7$
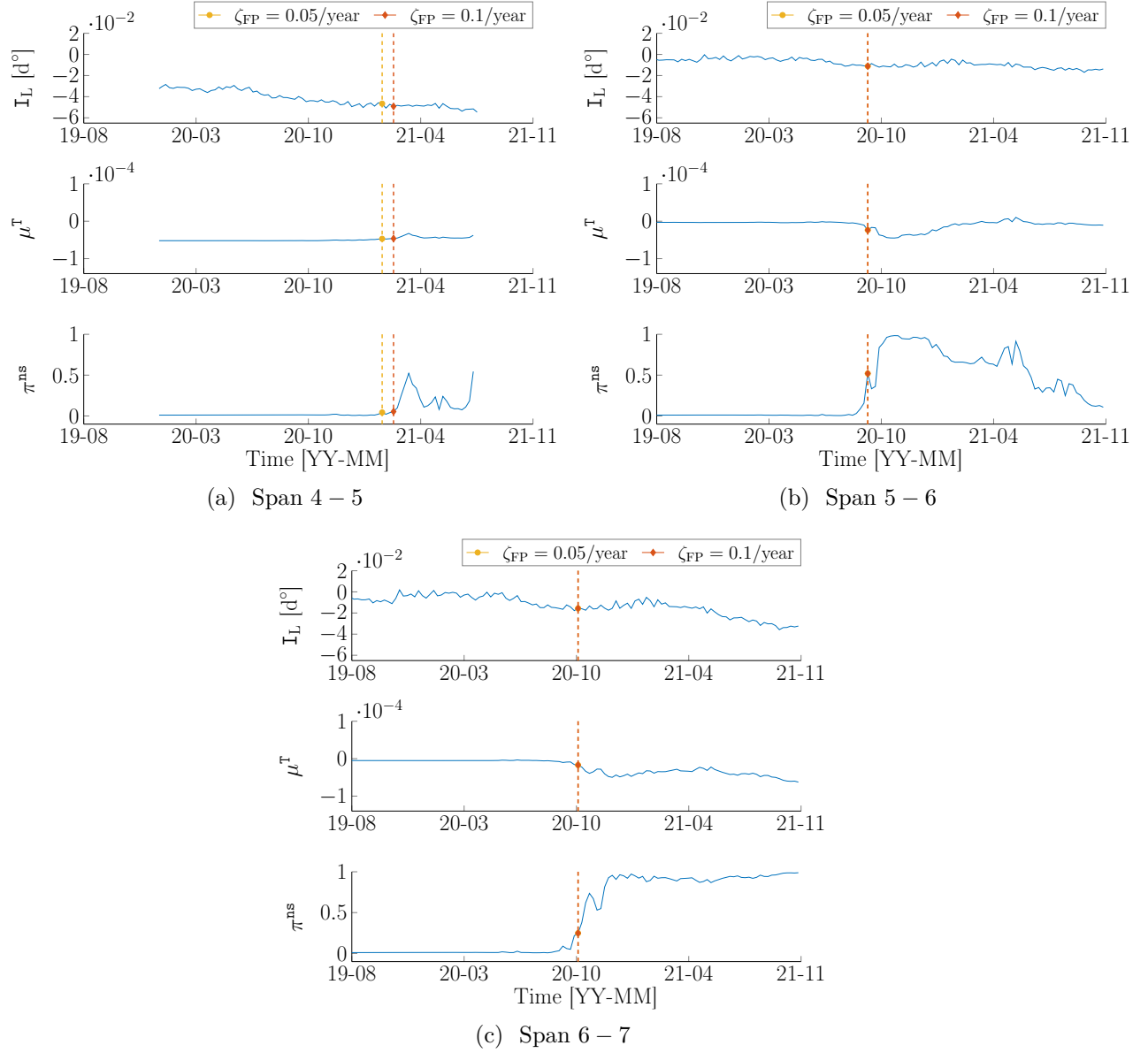
Figure 4.14 The anomaly detection for the collected lateral inclinations on axis 3, i.e., $\mathtt{I_L}$, across all the spans. The yellow marker corresponds to the annual false positive ratio $\zeta_{\mathrm{FP}} = 0.05/\mathrm{year}$, and the red one is associated with $\zeta_{\mathrm{FP}} = 0.10/\mathrm{year}$. For the span $4-5$, the agent associated with the $\zeta_{\mathrm{FP}} = 0.10/\mathrm{year}$ triggers an alarm 3 weeks later than the agent corresponding to the $\zeta_{\mathrm{FP}} = 0.05/\mathrm{year}$.

Such an expectation is met in cases like the ones shown in Figures 4.16b and 4.16c; These two figures are associated with the longitudinal inclinometers on the axis 3 corresponding to the spans $5-6$ and $6-7$ for which the agents with the annual false positive ratio of $\zeta_{\mathrm{FP}} = 0.10/\mathrm{year}$ trigger an alarm one week earlier than the agents associated with the annual false positive ratio of $\zeta_{\mathrm{FP}} = 0.05/\mathrm{year}$. Such a variability in detecting anomalies implies a need

for quantifying the true positive detections probabilistically as described in Section 4.2.3. Our results regarding the probability of true positive detections indicate that the agents with the annual false positive ratio $\zeta_{\text{FP}} = 0.1$/year detect anomalies earlier than the agents associated with $\zeta_{\text{FP}} = 0.05$/year. The detailed results and discussion for the probability of the true positive detections will be presented in Section 4.3.3.

Figure 4.15 illustrates the anomaly detection corresponding to the longitudinal inclinometers on the axis 1 across all instrumented spans, i.e., I_1. As shown, the agents associated with both annual false positive ration, i.e., $\zeta_{\text{FP}} = 0.05$/year and 0.10/year, result in the same timestamps when an alarm is triggered. Figure 4.16 illustrates the anomaly detection corresponding to the longitudinal inclinometers on the axis 3 across all instrumented spans, i.e., I_3. As shown, the agents associated with both annual false positive ratios, i.e., $\zeta_{\text{FP}} = 0.05$ and 0.10 per year, result in the same alarm triggering time for the span $4-5$. On the other hand, as mentioned earlier, this is not the case for the spans $5-6$ and $6-7$; The agent with a higher false positive ratio triggers an alarm one week later than the agents with the lower annual false positive ratio. Another point that needs to be mentioned is with regards to the occurrence of multiple anomalies. In some cases such as the ones shown in Figures 4.16b and 4.15b, there are multiple regime switches that the agents could trigger alarms for. However, the agents in this work are not suited to address multiple triggering, because they do not have a mechanism in order to continue evaluating the policies after the first time an alarm is triggered; As described in Sections 4.1 and 4.2.3, during the training phase, an agent-environment interaction is terminated whenever an agent takes the action of triggering an alarm.

4.3.3 Jacques Cartier Bridge Anomaly Detectability

We follow the procedure described in Section 4.2.3 along with the independent true positive test dataset $\mathcal{D}^{\text{test, TP}}$ presented in Table 4.4 in order to evaluate the probability of true positive detections for each agent. Figure 4.17 illustrates the probability of true detections associated with the sensors where I_L is the lateral inclinometer, I_1 is the longitudinal inclinometer 1, I_3 is the longitudinal inclinometer 3, and E is the extensometer. The three axes in each figure represent the covariates vector $\mathbf{x} = [|h_\text{a}| \; w_\text{a} \; \text{N}_{t>t_\text{s}}]^\intercal$ consisting in the absolute value of the anomaly height $|h_\text{a}|$, its duration w_a, and the number of time steps $\text{N}_{t>t_\text{s}}$ required for an agent to trigger an alarm after its starts.

Figure 4.17 illustrates the probability of true positive detections obtained for each sensor type, and for each target annual false positive ratio. The colors indicate the probability of true detections $\Pr(Y = +1 \equiv \text{true positives}|\mathbf{x}) \in (0, 1)$. In addition, Figures 4.17a-4.17d show the the probability of true detections for the annual false positive $\zeta_{\text{FP}} = 0.05$/year,
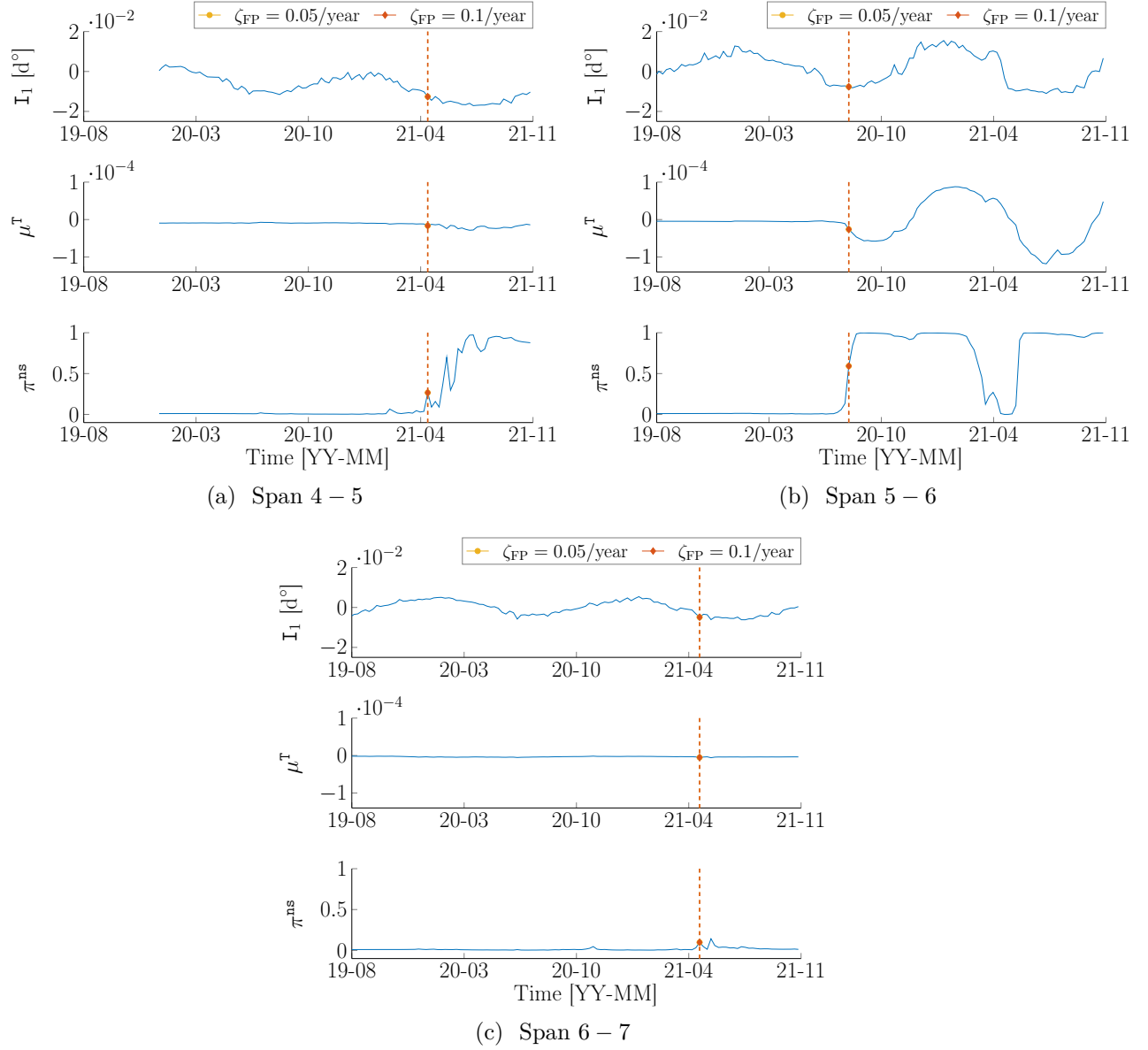
Figure 4.15 The anomaly detection for the collected longitudinal inclinations on axis 1, i.e., I_1, across all the spans. The yellow marker corresponds to the annual false positive ratio $\zeta_{FP} = 0.05$/year, and the red one is associated with $\zeta_{FP} = 0.10$/year.

while Figures 4.17e-4.17h are associated with the $\zeta_{FP} = 0.1$/year.

Figure 4.18 shows the isosurfaces corresponding to a 90% probability of true positive detections extracted from the results obtained by TAGI and shown in Figure 4.17. The blue plane quantifies the minimum detectable anomaly magnitude $|h_a|_{\min}$. The yellow plane quantifies the minimum detectable anomaly growth rate per year $\psi_{\min} = \psi_0 \times |h_a|_{\min}$ as a function of its corresponding $|h_a|_{\min}$.
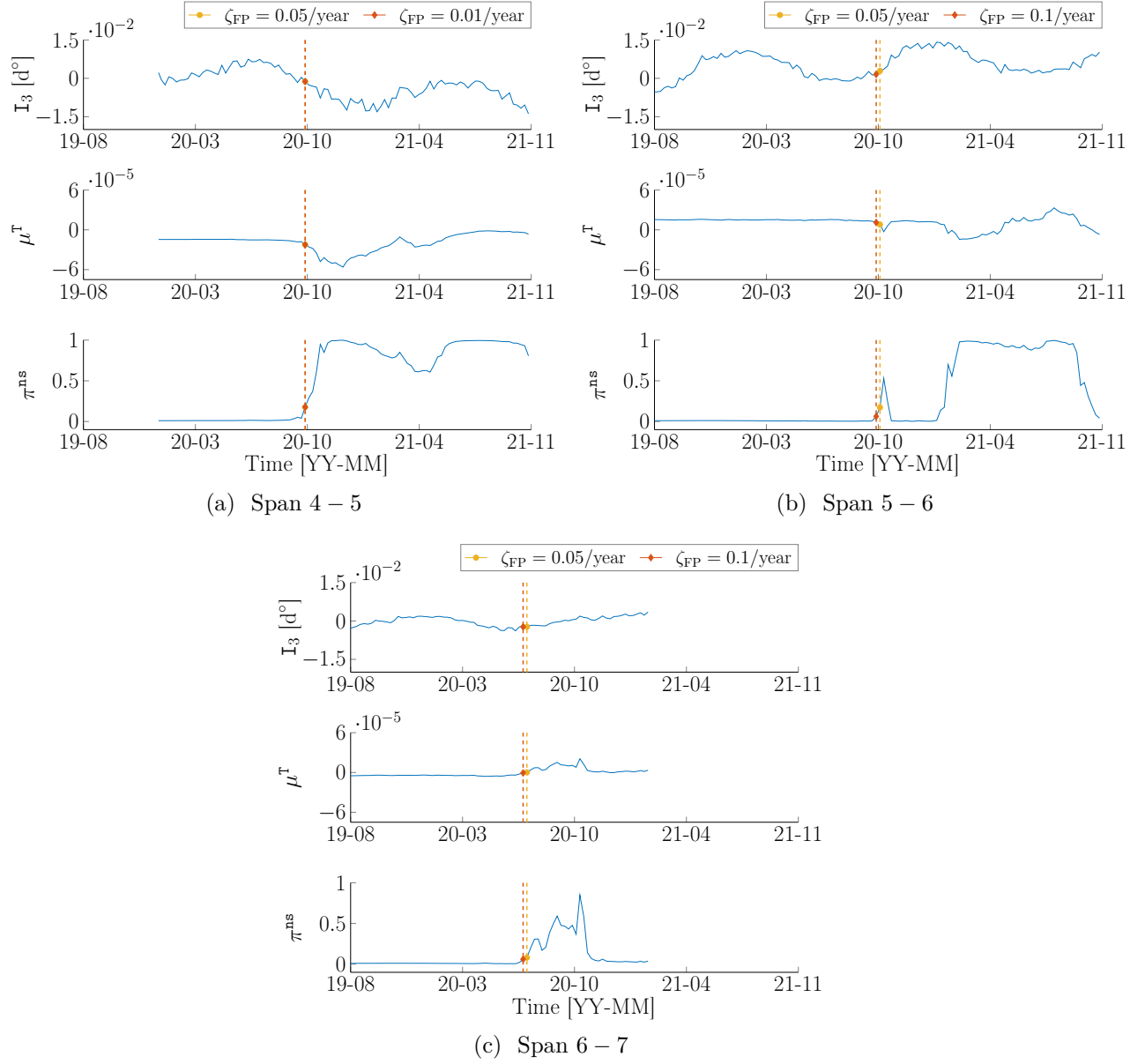
Figure 4.16 The anomaly detection for the collected longitudinal inclinations on axis 3, i.e., I_3, across all the spans. The yellow marker corresponds to the annual false positive ratio $\zeta_{FP} = 0.05$/year, and the red one is associated with $\zeta_{FP} = 0.10$/year. For the span $5-6$ and $6-7$, the agents associated with the $\zeta_{FP} = 0.10$/year triggers an alarm 1 week earlier than the agent corresponding to the $\zeta_{FP} = 0.05$/year.

For a better illustration, we re-draw Figure 4.18 from top view so it shows the figures form the $|h_a| - w_a$ plane perspective. Figure 4.19 presents the results and the induced hue effects are caused by such a rotation. Figures 4.19a-4.19d correspond to an annual false positive ratio $\zeta_{FP} = 0.05$/year, while Figures 4.19e-4.19h are associated with the $\zeta_{FP} = 0.10$/year.
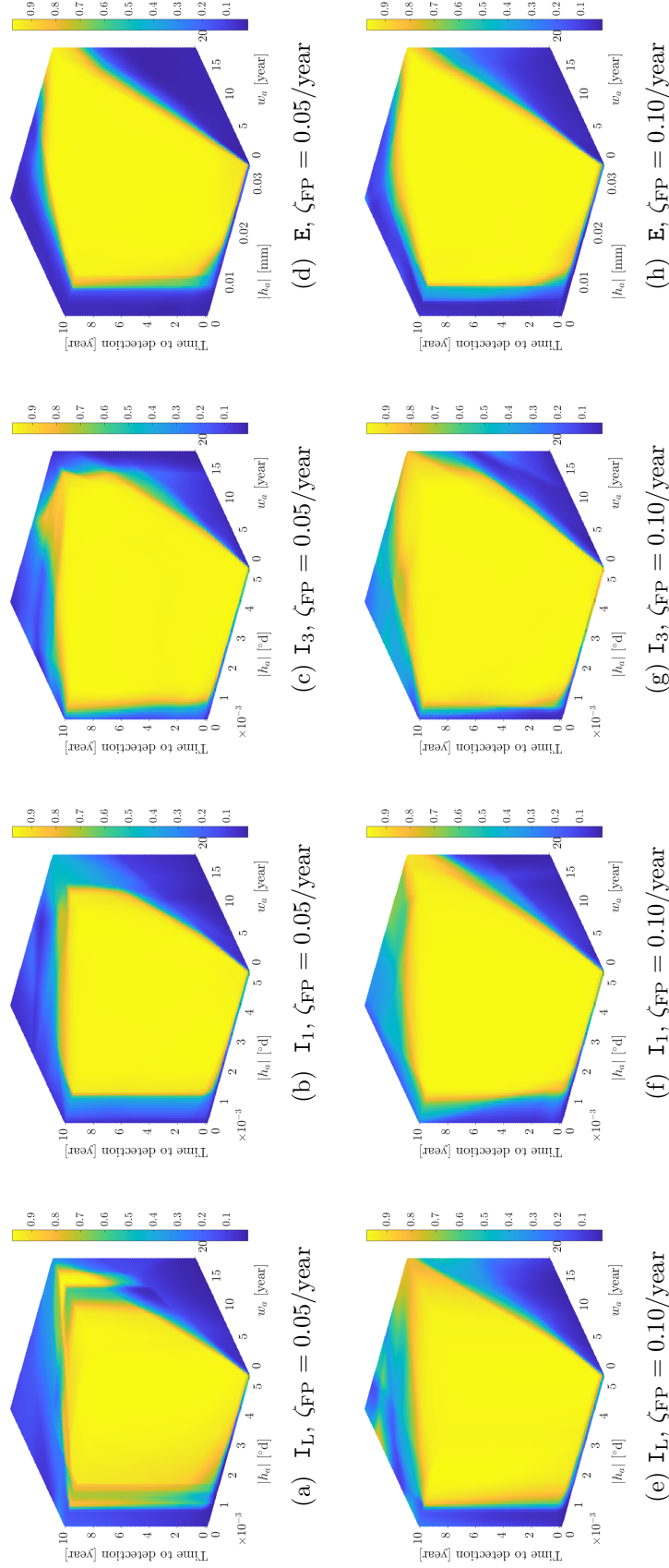
Figure 4.17 Illustration of the probability of the true positive detection for: (a-d) $\zeta_{FP} = 0.05$, and (e-h) $\zeta_{FP} = 0.1$/year. The sensor names are I_L: Lateral inclinometer, I_1: Longitudinal inclinometer 1, I_3: Longitudinal inclinometer 3, and E: Extensometer. In addition, the unit for the anomaly magnitude depends on the sensor type, that is mm for the extensometer and °d for the inclinometers.
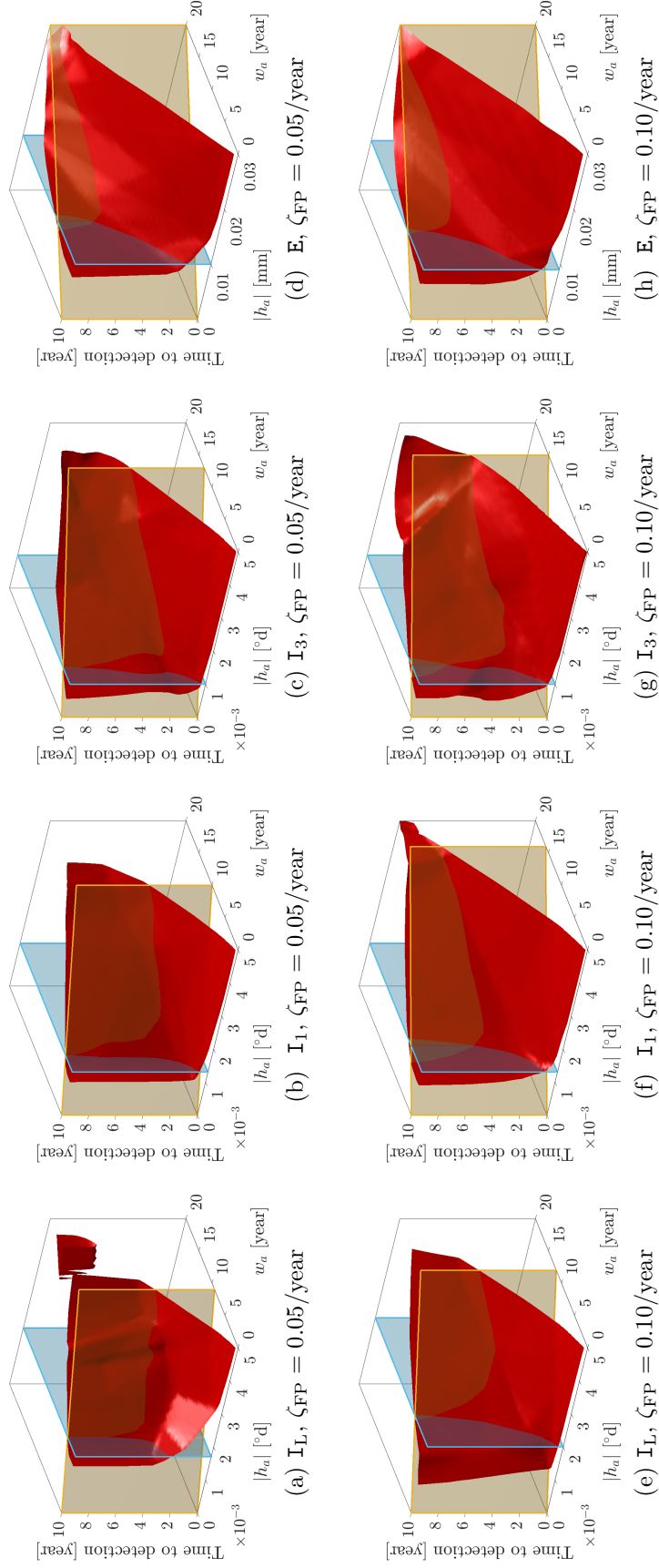
Figure 4.18 Illustration of the 90% probability for true positive detections for: (a-d) $\zeta_{\mathrm{FP}} = 0.05$, and (e-h) $\zeta_{\mathrm{FP}} = 0.1$/year. The sensor names are $\mathrm{I_L}$: Lateral inclinometer, $\mathrm{I_1}$: Longitudinal inclinometer 1, $\mathrm{I_3}$: Longitudinal inclinometer 3, and E: Extensometer. The blue plane quantifies the minimum detectable anomaly magnitude $|h_a|_{\mathrm{min}}$. The yellow plane quantifies the minimum detectable anomaly growth rate per year $\psi_{\mathrm{min}} = \psi_0 \times |h_a|_{\mathrm{min}}$ as a function of its corresponding $|h_a|_{\mathrm{min}}$.

(a) I_L, $\zeta_{FP} = 0.1/year$

(b) I_1, $\zeta_{FP} = 0.05/year$

(c) I_3, $\zeta_{FP} = 0.05/year$

(d) E, $\zeta_{FP} = 0.05/year$

(e) I_L, $\zeta_{FP} = 0.1/year$

(f) I_1, $\zeta_{FP} = 0.1/year$

(g) I_3, $\zeta_{FP} = 0.1/year$
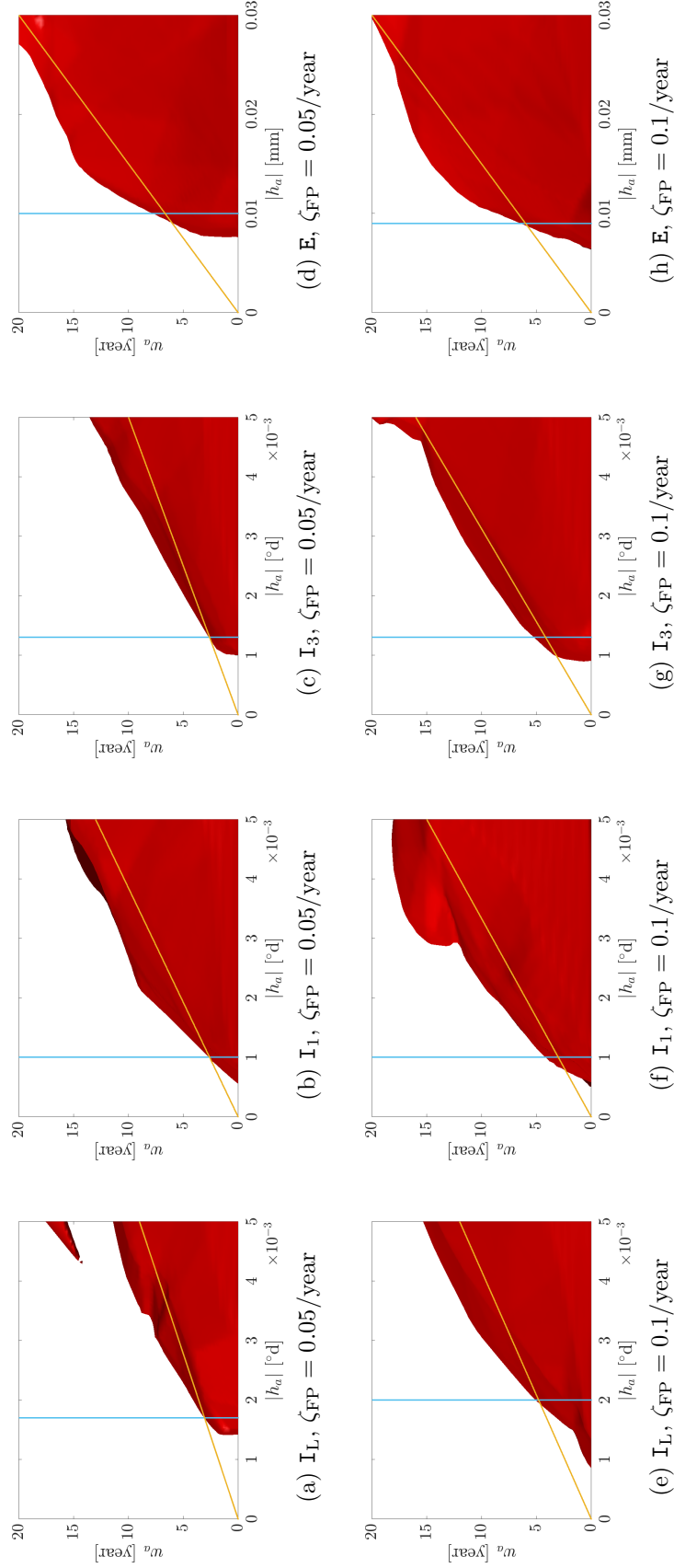
(h) E, $\zeta_{FP} = 0.1/year$

Figure 4.19 Comparison of the isosurface corresponding to 90% probability of true positive detections in $|h_a| - w_a$ plane for: (a-d) $\zeta_{FP} = 0.05$, and (e-h) $\zeta_{FP} = 0.1/year$. The sensor names are I_L: Lateral inclinometer, I_1: Longitudinal inclinometer 1, I_3: Longitudinal inclinometer 3, and E: Extensometer. The blue line indicates the minimum detectable anomaly magnitude $|h_a|_{min}$ with the probability of 90%. The red line indicates the minimum detectable anomaly growth rate ψ per year with the same probability of 90%.

The comparison between these two cases confirms that increasing the annual false positive ratio results in an increase in the true positive detections. Such a tradeoff is seen from two comparisons: (i) comparing the time to detect an anomaly; For an anomaly with a fixed magnitude and duration, Figures 4.19e-4.19h which correspond to a higher annual false positive ratio, show earlier anomaly detections as opposed to Figures 4.19a-4.19d which are associated with a lower ζ_{FP}, and (ii) comparing the red area between the two annual false positive ratios ζ_{FP}; For the higher annual false positive ratio, the red areas increase in both directions with respect to the anomaly magnitude and duration. This implies that for the case with $\zeta_{FP} = 0.10$/year, the agent is able to detect anomalies with lower magnitudes and with longer durations. Note that the planes in Figure 4.18 are selected such that: (i) the area enclosed by the planes creates a volume for which all the data points corresponding to the probability of the true detections $\Pr(Y = +1 : \mathrm{TP}|\mathbf{x}) \geq 90\%$, and (ii) the yellow planes, which correspond to the anomaly growth rate, pass the origin, so the interpretation of the ψ is simplified to a single linear function, i.e., $\psi_{\min} = \psi_0 \times |h_a|_{\min}$. Nevertheless, for specific needs, a more complex function could be used in order to represent more precisely the minimum anomaly growth rate.

Table 4.7 summarizes the minimum detectable anomaly magnitude and its growth rate. The first column shows the shorthand and complete name of the sensors. The second column is associated with the agents with the annual false positive ratios of $\zeta_{FP} = 0.05$ and 0.1 per year. The third column identifies the minimum detectable anomaly magnitude $|h_a|_{\min}$, and the fourth one shows the the constant ψ_0 corresponding to the minimum detectable anomaly growth rate $\psi_{\min}$ per year. The last column shows the minimum detectable anomaly growth rates $\psi_{\min}$ in years as a function of $|h_a|_{\min}$, i.e., $\psi_{\min} = |h_a|_{\min}$ per $\mathtt{N}_\psi$ years. For instance,

Table 4.7 Minimum detectable anomaly magnitudes $|h_a|_{\min}$ and the the constant ψ_0 corresponding to the minimum detectable anomaly growth rates in year such that $\psi_{\min} = \psi_0 \times |h_a|_{\min}$. The last column show the minimum detectable anomaly growth rates $\psi_{\min}$ in years as a function of $|h_a|_{\min}$, i.e., $\psi_{\min} = |h_a|_{\min}$ per $\mathtt{N}_\psi$ years.

Sensor name	ζ_{FP}	$|h_a|_{\min}$	ψ_0	$\mathtt{N}_\psi$ [year]
$\mathtt{I_L}$: Lateral inclinometer	0.05	0.0017 °d	0.15	7
	0.10	0.0020 °d	0.13	8
$\mathtt{I_1}$: Longitudinal inclinometer 1	0.05	0.0013 °d	0.19	6
	0.10	0.0013 °d	0.19	6
$\mathtt{I_3}$: Longitudinal inclinometer 3	0.05	0.0010 °d	0.25	4
	0.10	0.0010 °d	0.25	4
$\mathtt{E}$: Extensometer	0.05	0.0100 mm	0.15	7
	0.10	0.0090 mm	0.17	6

for the agent corresponding to the lateral inclinometer with the annual false alarm ratio of $\zeta_{\text{FP}} = 0.05/\text{year}$, the minimum anomaly magnitude must be equal or greater than $0.0017°\text{d}$ per $\mathbb{N}_\psi = 7$ years. In other words, an anomaly with a magnitude of $0.0017°\text{d}$ is detectable, if it develops in less than seven years.

4.4   Discussion

The goal of reinforcement learning is to determine optimal policies by maximizing the accumulated discounted rewards. While succeeding in the anomaly detection problem, reinforcement learning suffers from two limitations with respect to the reward function. The first limitation is related to designing a reward function in order to accurately achieve a desired behavior [88]. In this regard, challenges are associated with the designer's understanding about the task in hand, and representing the task in the form of a reward function [89, 90]. In other words, designing an appropriate reward function relies on hand-crafted human expertise; In the context of anomaly detection, Section 4.1.2 introduced a reward function that quantifies the reward values based on the time when an alarm is triggered with respect to the anomaly's start time. From the decision-maker's perspective, an alarm should be triggered right after an anomaly starts, while preventing false alarms when no anomaly is present. In addition, the time of detection for an anomaly should be as close as possible to the time when the anomaly starts, so it prevents a delayed triggering points or missed alarms. Such desired behavior results in the positive and negative reward values in order to either encourage or prevent an action taken by the agent.

The second limitation is associated with the reward sparsity problem [91]; It refers to a situation when a reward function only encourages or prevents an agent for taking an action in limited situations. For instance, if we only consider a single positive reward value for triggering an alarm after it starts, an agent needs to explore more in order to learn the optimal policy, because there is not a feedback signal available for the agent with respect to false or missed detections. Therefore, the reward sparsity affects the cost of the learning process with respect to the amount of agent-environment interactions required during the learning process. Reward functions such as the one presented in Section 4.1.2 relaxes this limitation by using additional reward features related to the actions taken by the agent in different situations [92]; For the anomaly detection problem, it involves assigning different reward values that accounts for different time to triggering an alarm as it is shown in Section 4.1.2. Note that the reward sparsity limitation is less relevant when generating episodes is cheap such as the simulation process presented in Section 4.1.1 where a generic probabilistic model is used to generate abnormal episodes.

4.5   Conclusion

This chapter introduced a framework in order to detect anomalies using reinforcement learning (RL). The anomaly detection problem was formulated as the identification of alarm triggering policies in the context of the RL. In this regard, this chapter introduced methodologies in order to build environments that the agents interact with; Each environment was built using episodes that consist in the structural degradation's speed and the probability of a non-stationary regime. Each episode was obtained from a generic time series simulation procedure that involves using empirical models of the structural responses as well as a stochastic anomaly function described in Section 4.1.1. As a result, all the agents were able to be trained with a wide range of anomalies. Quantifying the anomalies with respect to their characteristics is a crucial aspect in detecting anomalies because while we do not know what are the durations and magnitudes of the anomalies as well as their time of occurrence, we are interested in the anomalies with low magnitudes that develop over several years.

Furthermore, this chapter presented methodologies in order to quantify the detectability of anomalies for each sensor type. To this end, two metrics were introduced which represent the annual false positive ratio and the probability of true positive detections. It was found that the reward function values control the behavior of the agents with respect to the number of false alarms per year. This is an important step in the anomaly detection procedure in the sense that the lack of control over the annual false alarm ratios could result in an unmanageable number of wrongful anomaly detections, which subsequently could bring about additional costs. Moreover, this chapter examined the probability of true positive detection for the agents with different false positive ratios. This examination resulted in isosurfaces identifying the probability of a true positive detection with respect to an anomaly's magnitude and its duration, as well as the number of time steps an agent takes in order to detect an anomaly after its start. Results have confirmed empirically that for a higher annual false positive ratio, a 90% probability of true positive detections is achieved for an anomaly with lower magnitude and longer duration. Moreover, this chapter summarized the detectability of each sensor type with respect to two macroscopic quantities representing the minimum detectable anomaly magnitude and its growth rate. These two quantities give simple and interpretable information about the anomaly detectability of each sensor type.

Moreover, this chapter discussed the limitations with respect to the RL-based anomaly detection; They are: (i) designing a hand-crafted reward values such that they are able to encourage ro prevent an action taken by the agent, and (ii) preventing the reward sparsity by designing different reward values that allows an agent receiving feedback signals for different situations. Before presenting the methodologies developed for associating the severity

levels of structural damages to the magnitudes of the anomalies, Chapter 5 will present an alternative approach to address the limitations regarding the reward function by allowing the agent to learn alarm triggering policies without having access to a reward function.

CHAPTER 5   Anomaly Detectability Using Imitation Learning

This chapter presents a method relying on the imitation learning approach for detecting anomalies and for quantifying the detectability of anomalies. To this end, the objective of an agent is to learn to imitate a *demonstration set* obtained from an expert's policy. In this regard, we present the procedure of establishing a demonstration set based on an expert's decision with respect to triggering an alarm for different time steps during an episode. Finally, we use the methodologies developed in Chapter 4 for quantifying the anomaly detectability on the measurements collected on the three instrumented spans of the Jacques Cartier Bridge. Such a quantification leads to the same probabilistic metrics presented in Section 4.2 which are the annual false detection ratio, and the probability of true positive detections.

5.1   Anomaly Detection Using Imitation Learning

The main difference between a reinforcement learning agent and an imitation learning one is that the former uses a reward function as a feedback signal to identify the optimal policy, while the latter uses a demonstration set established by an expert in order to imitate the expert's policy. We can see the imitation learning approach as a classification problem in the context of the supervised learning approach where the labels are provided by an expert. A demonstration is represented by an episode $\hat{\mathcal{Z}}$ that is defined as a set of state-action pairs for consecutive time steps following

$$\hat{\mathcal{Z}} = \{(\mathbf{s}_1, a_1^{\text{expert}}), (\mathbf{s}_2, a_2^{\text{expert}}), \cdots\}, \tag{5.1}$$

where, at a time t, $\mathbf{s}_t = [\mu_t^{\mathsf{T}} \ \pi_t^{\mathsf{ns}}]^{\mathsf{T}}$ is the environment's state vector, and $a_t^{\text{expert}} = \pi^{\text{expert}}(\mathbf{s})$ is the action obtained from the expert policy. An agent uses a set of expert demonstrations $\Xi = \{\hat{\mathcal{Z}}_1, \hat{\mathcal{Z}}_2, \cdots, \hat{\mathcal{Z}}_{\mathtt{Z}}\}$ consisting in $\mathtt{Z}$ demonstration sets in order to learn a policy $\pi(\mathbf{s})$ that imitates the expert's policy $\pi^{\text{expert}}(\mathbf{s})$. In this work, an agent directly learns to imitate the expert's policy using the classification approach discussed in Section 2.3.2. Therefore, the optimal policy $\pi^*(\mathbf{s})$ is obtained by building a classifier for a set of $\mathtt{D}$ joint observations $\mathcal{D} = \{(\mathbf{s}_i, a_i^{\text{expert}}), \forall i \in \{1 : \mathtt{D}\}\}$, where, $\mathbf{s}_i = [\mu_t^{\mathsf{T}} \ \pi_t^{\mathsf{ns}}]^{\mathsf{T}}$ are the covariates vector and $a_i^{\text{expert}} \in \{-1 : a^0 \equiv \text{trigger an alarm}, +1 : a^1 \equiv \text{not trigger an alarm}\}$ are the system responses representing the actions taken by the expert. Here, we employ the *tractable approximate Gaussian inference* (TAGI) method for Bayesian neural networks in order to build a classification model [87] from demonstration sets. For the problem at hand, the proba-

bility of the two classes are: (i) the probability of triggering an alarm $\Pr(A = +1 : a^0|\mathbf{s})$, conditional on the covariates vector $\mathbf{s}$, and (ii) the probability of not triggering an alarm $\Pr(A = -1 : a^1|\mathbf{s}) = 1 - \Pr(A = +1 : a^0|\mathbf{s})$.

Furthermore, in order to quantify the anomaly detectability of each agent, we carry out the same procedure presented in Section 4.2; This implies that we apply the classifier built on two independent test datasets, i.e., $\mathcal{D}^{\text{test, FP}}$ and $\mathcal{D}^{\text{test, TP}}$, in order to quantify the annual false positive ratio and the probability of true positive detections. Here, we use the same datasets that were used for quantifying the anomaly detectability using reinforcement learning; the superscripts $^{\text{test, FP}}$ and $^{\text{test, TP}}$ are associated with the datasets built for evaluating the false positive and true positive detections, respectively. The characteristics of the datasets are presented in Table 4.4 in Chapter 4.

5.2 Demonstration Set Based on an Expert's Policy

In order to build a demonstration set Ξ, we choose $\mathbb{Z}$ episodes from the training data set $\mathcal{D}^{\text{train}}$ presented in Table 4.4 for each sensor. For each episode, we assume that (i) the expert policy $\pi^{\text{expert}}(\mathbf{s})$ for a visited state $\mathbf{s}$ is to trigger an alarm for the time steps during the anomaly window defined in Section 4.1.1 and not to trigger an alarm for remaining time steps, and (ii) the expert policy $\pi^{\text{expert}}(\neg\mathbf{s})$ for a non-visited or out-of-domain state $\neg\mathbf{s}$ is to trigger an alarm for all time steps. Therefore, the expert policy is

$$\pi^{\text{expert}}(\mathbf{s}) = \begin{cases} a^0, & \forall t \in (t_s, t_e) \\ a^1, & \text{Otherwise} \end{cases}, \quad \pi^{\text{expert}}(\neg\mathbf{s}) = a^0, \forall t, \tag{5.2}$$

where, t_s and t_e are the anomaly's start and end time, respectively. Figure 5.1 shows an example of an abnormal episode used by an expert in order to build a demonstration according to Equation 5.2. Figure 5.1a illustrates the underlying anomaly function simulated in order to generate the synthetic time series shown in Figure 5.1c. The two vertical dashed lines indicate the beginning and the end of the anomaly for which its magnitude $h_{\text{a}} = 0.011$ unit and its duration w_{a} is a year. For a better illustration, the anomaly duration is indicated by a shaded area. Note that the units of the anomaly function is the same as the units of the simulated time series. Figures 5.1b and 5.1d show the trend expected value μ^{T} and the probability of the non-stationary regime π^{ns} obtained from the BDLM analysis of the simulated time series. The blue and red lines in Figure 5.1 indicate the normal and abnormal time steps, respectively. According to the expert's decision presented in Equation 5.2, the action for the normal time steps is to not trigger an alarm, while the action for the abnormal

(a) Anomaly fucntion

(b) Expected value of the local level

(c) Simulated time series

(d) Probability of the non-stationary regime
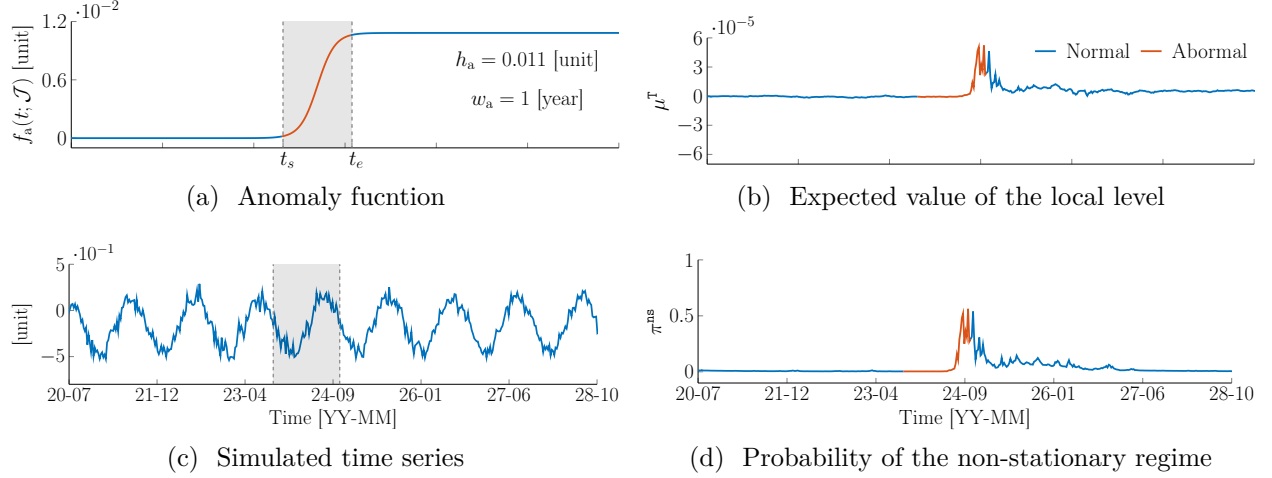
Figure 5.1 Example of an abnormal episode used by an expert in order to build a demonstration. The realization of the anomaly function results in an anomaly magnitude of $h_a = 0.11$ and a duration of $w_a = 1$ year. The shaded area represents the anomaly duration with the start t_s and end t_e times. The red and blue lines indicate the abnormal and normal time steps, respectively. Their respective actions according to Equation 5.2 are a^0 : trigger an alarm, and a^1 : not trigger an alarm.

time steps is to trigger an alarm. Therefore, for this time series, from a total number of 522 data points, the demonstration set consists in the 56 state-action pairs associated with the action of triggering an alarm, and 466 ones for the action of not triggering an alarm.

In Equation 5.2, the action of triggering an alarm is only assigned for the time steps associated with the anomaly window, e.g., the shaded area in Figure 5.1a which corresponds to the time steps $t \in (t_s, t_e)$. This consideration is due to the fact that the anomaly function for $t \notin (t_s, t_e)$ is constant and it results in a zero derivative for these values. This implies that the structural degradation's speed is zero for these time steps and therefore, the optimal policy is to not trigger an alarm. Note that, in the reinforcement learning approach we let an agent to make the decision of triggering an alarm by relying on the reward values. Therefore, an agent might trigger an alarm with a delay, even trigger an alarm after an anomaly ends. However, in the imitation learning approach, we ensure that an agent follows the expert's policy.

We employ the aforementioned procedure on simulated time series with a wide range of anomalies in order to establish a training dataset and build a classifier. Before training an agent to imitate the expert's policy, we reduce the number of samples in order to accelerate the training procedure. To this end, first we use a grid $d_{M \times N}$, and apply it on the environment's state-space in order to create $M \times N$ unique regions. An example of such regions is shown in Figure 4.6 where a grid of 20×20 results in 400 unique regions. Afterwards, wherever

it is possible, we randomly draw 10 samples from each of these regions. This approach enables building a training dataset for which the demonstrations are collected from the whole environment's domain, while reducing the number of samples. The grid characteristics, i.e., `M` and `N`, are presented in Table 4.5. As it will be shown in Section 5.3, such downsampling allows the agent to learn the optimal policy with less episodes and consequently at a lower computational expense. This was already noted in the work of Syed and Schapire [93], where they proved that imitation learning requires fewer demonstrations than reinforcement learning in order to imitate an expert. Moreover, we do not use the classifier built in order to extrapolate the actions predicted for the non-visited or out-of-domain states. Instead, we force the agent to imitate the expert's decision for these states that is to trigger an alarm. To this end, we add a third class for such states when building the classifier, and later directly assign the action of triggering an alarm for these states. Such non-visited states are shown by a white color in Figure 4.6. In order to generate data for this new class, we draw one sample from each non-visited cell. This preprocessing coincides with the actions selected for the non-visited or out-of-domain states when using the reinforcement learning approach presented in Chapter 4.

5.3 Results

In this section, we first present the demonstration set Ξ used for the imitation learning approach and subsequently for building a classifier in order to detect anomalies on the three instrumented spans of the Jacques Cartier Bridge. Afterwards, we present the anomaly detectability of the classifier with respect to the annual false positive ratio as well as the probability of true positive detections.

5.3.1 Jacques Cartier Bridge Training and Test Demonstration Sets

We choose `Z` = 100000 episodes from the training data set $\mathcal{D}^{\text{train}}$ for each sensor. This results in approximately 34 million data points representing the visited states. We follow the preprocessing procedures explained in Section 5.2 in order to reduce the number of samples and add samples for the third class corresponding to the non-visited states. This results in a dataset with a total of approximately 33000 samples, which is further splitted into a training and test sets such that 70% of the samples are used to train a classifier. Note that the number of episodes we used to train an agent using the imitation learning approach is approximately 8% of the number of episodes that were used to train an agent using the reinforcement learning approach.

Figure 5.2 shows the resulting demonstration sets for the inclinometers and extensometers. Figures 5.2a-5.2d correspond to the training demonstration sets, while Figures 5.2e-5.2h are associated with the test ones. Note that the process of reducing the number of samples for the training and test demonstration sets remain the same. The red dots are associated with the abnormal states which correspond to the time steps during an anomaly.
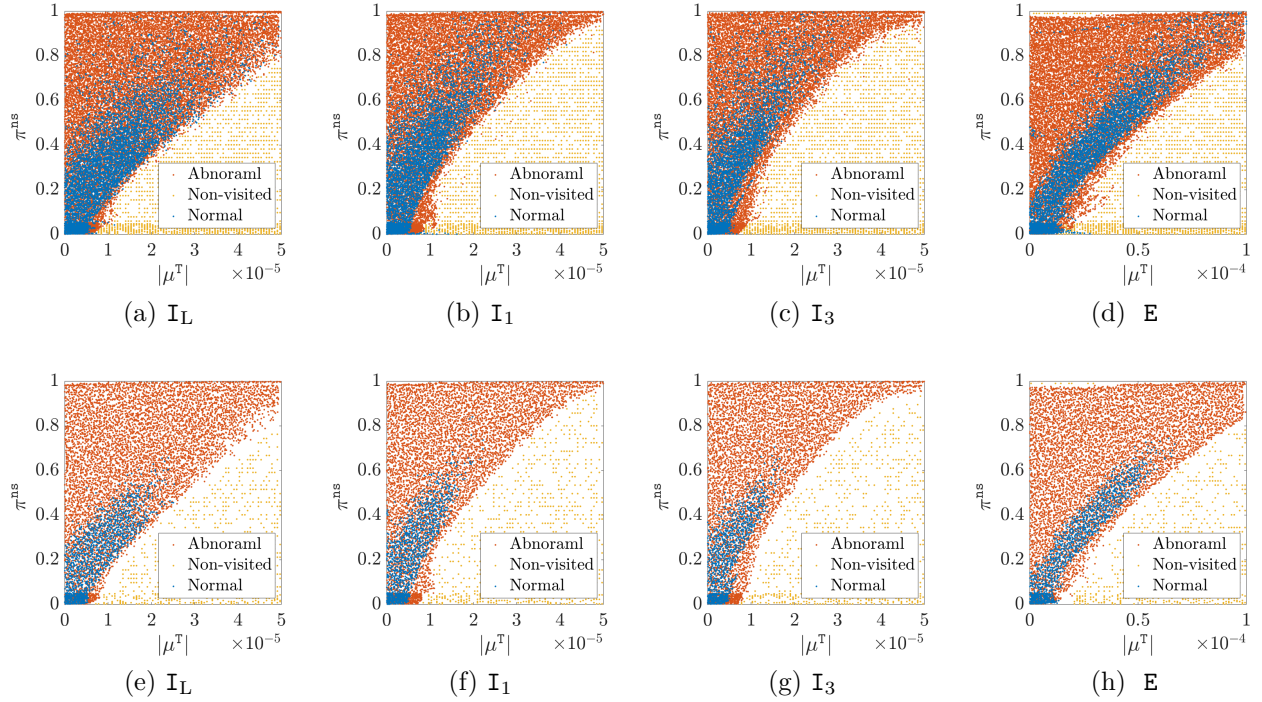
Figure 5.2 Illustrations of (a-d) the training demonstration sets and (e-h) the test demonstration sets for inclinometers and extensometers. The red dots are associated with the abnormal visited states corresponding to the time steps during the anomaly for which the expert's decision is a^0 : trigger an alarm. The blue dots are associated with the normal visited states corresponding to the remaining time steps for which the expert's decision is a^1 : not trigger an alarm. The yellow dots represent the non-visited states for which the agent is forced to take the action of triggering an alarm by imitate the expert's decision. The sensor names are I_L: Lateral inclinometer, I_1: Longitudinal inclinometer 1, I_3: Longitudinal inclinometer 3, and E: Extensometer.

5.3.2 Jacques Cartier Bridge Anomaly Detection

We use TAGI [87] in order to build a classifier for each sensor's demonstration set as presented in Section 5.1. Figure 5.3 illustrates the policy of each agent for the inclinometers and extensometers. Applying the policies shown in Figure 5.3 on the corresponding false positive test datasets $\mathcal{D}^{\text{test},\text{FP}}$ results in different annual false positive ratios ζ_{FP}, that is 0.008 false

$$\text{(a) } \mathtt{I_L} \qquad\qquad \text{(b) } \mathtt{I_1} \qquad\qquad \text{(c) } \mathtt{I_3} \qquad\qquad \text{(d) } \mathtt{E}$$
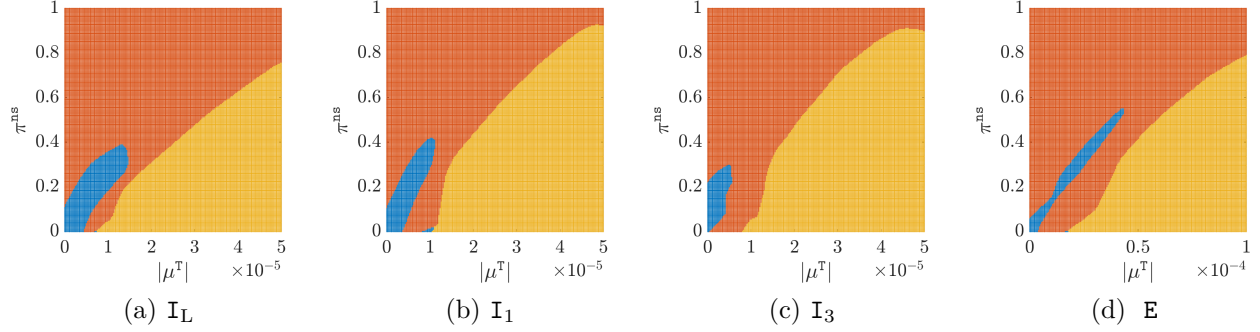
Figure 5.3 The policy maps for the inclinometers and the extensometers. The boundaries are obtained for the probability of triggering alarms $\Pr(Y = -1 : a^0|\mathbf{s}) > 0.5$. The red color indicates that the action is a^0: trigger an alarm, and the blue one indicates that the action is a^1: not trigger an alarm. The yellow areas are associated with the non-visited states for which the action is a^0: Trigger an alarm. Also, $\mathtt{I_L}$: Lateral inclinometer, $\mathtt{I_1}$: Longitudinal inclinometer 1, $\mathtt{I_3}$: Longitudinal inclinometer 3, and $\mathtt{E}$: Extensometer.

positive per year for the lateral inclinometer, 0.006 false positive per year for the longitudinal inclinometer 1, 0.2 false positive per year for the longitudinal inclinometer 3, and 0.028 false positive per year for the extensometer. Such an evaluation of the annual false positive ratios in Figure 5.3 is not acceptable because we are interested in having control over the number of false positives per year for each sensor. To overcome this issue, we applied a grid search on an independent test dataset over the alarm triggering probabilities in order to obtain the target annual false positive ratios.

Table 5.1 summarizes the minimum probability of triggering an alarm $\Pr_{\min}(A = -1 : a^0|\mathbf{s})$ for the inclinometers and extensometers in order to obtain one false positive detection every 5, 10, 20, and 40 years. These values are equivalent to annual false positive ratios ζ_{FP} of 0.2, 0.1, 0.05, and 0.025 false alarms per year. As expected, the probability val-

Table 5.1 The minimum probability of triggering an alarm $\Pr_{\min}(A = -1 : a^0|\mathbf{s})$ for the annual false positive ratios ζ_{FP} of 0.2, 0.1, 0.05, and 0.025 false alarms per year.

Sensor name	ζ_{FP}			
	0.2	0.1	0.05	0.025
$\mathtt{I_L}$: Lateral inclinometer	0.48017	0.48019	0.48023	0.48055
$\mathtt{I_1}$: Longitudinal inclinometer 1	0.47655	0.47700	0.47770	0.47880
$\mathtt{I_3}$: Longitudinal inclinometer 3	0.50000	0.50350	0.50630	0.51900
$\mathtt{E}$: Extensometer	0.47608	0.48100	0.49000	0.50300

ues for triggering alarms monotonically increase as the target annual false positive ratio decreases. Figure 5.4 illustrates the contours for the minimum probability of triggering an alarm $\Pr_{\min}(Y = -1 : a^0|\mathbf{s})$ associated with Table 5.1 for the inclinometers and extensometers. This particular contour line represents the policy maps presented in Figure 5.3. For a better illustration, Figures 5.4e-5.4h correspond to the contours from Figures 5.4a-5.4d, but for a restricted domain with respect to the probability of regime switching, i.e., $\pi_t^{\mathbf{ns}} \in (0, 0.6)$. In Figure 5.4, the spacing between some contour lines are so small that in some cases they are not distinguishable, e.g., Figure 5.4e and Figure 5.4f that correspond to the lateral inclinometer and longitudinal inclinometer 1, respectively. This shows the sensitivity of the imitation learning approach for smaller annual false positive ratios. Therefore, controlling the number of false positives is difficult in the sense that an engineer needs to try different

(a) $\mathtt{I_L}$ (b) $\mathtt{I_1}$ (c) $\mathtt{I_3}$ (d) $\mathtt{E}$

(e) $\mathtt{I_L}$ (f) $\mathtt{I_1}$ (g) $\mathtt{I_3}$ (h) $\mathtt{E}$
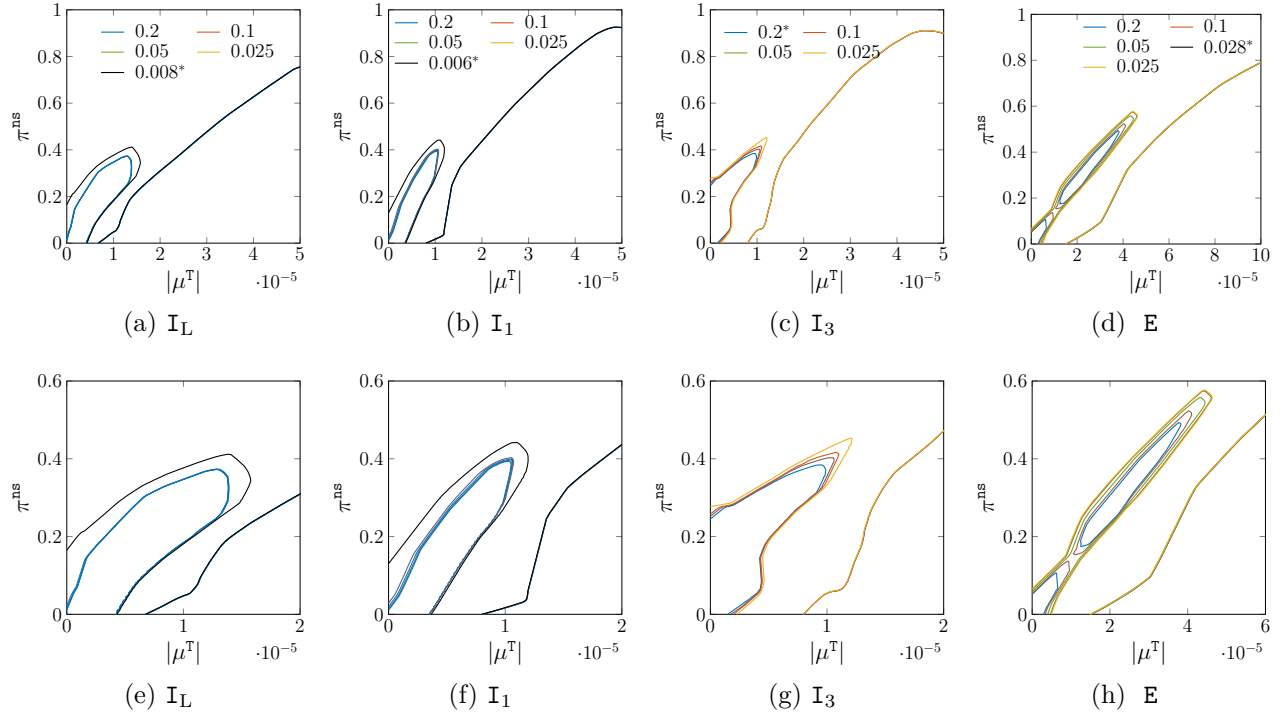
Figure 5.4 The policy map contours for the inclinometers and the extensometers: (a-d) the policy maps for the probability of the regime switching $\pi_t^{\mathbf{ns}} \in (0, 1)$, and (a-d) the policy maps for the probability of the regime switching $\pi_t^{\mathbf{ns}} \in (0, 0.6)$. The contour lines represent the annual false positive ratios ζ_{FP} of 0.2, 0.1, 0.05, and 0.025. The corresponding probabilities of triggering an alarm $\Pr(Y = -1 : a^0|\mathbf{s})$ are provided in Table 5.1. The black contour lines indicate the probability of triggering alarms $\Pr(Y = -1 : a^0|\mathbf{s}) > 0.5$ as shown in Figure 5.3. The corresponding annual false positive ratios are indicated by a superscript $*$. Also, $\mathtt{I_L}$: Lateral inclinometer, $\mathtt{I_1}$: Longitudinal inclinometer 1, $\mathtt{I_3}$: Longitudinal inclinometer 3, and $\mathtt{E}$: Extensometer.

probability values with high accuracy in order to achieve a target annual false positive ratio. On the other hand, reinforcement learning-based anomaly detection relies on selecting a reward values set and train an agent accordingly. The reinforcement learning approach results in a more robust approach in order to control the annual false positive ratio as shown in Figure 4.13; In this figure, the alarm triggering policies associated with different annual false positive ratios are distinguishable: The blue regions correspond to the action of not triggering an alarm that are recognizable for the annual false positive ratios ζ_{FP} of 0.1 and 0.05 per year.

5.3.3   Jacques Cartier Bridge Anomaly Detectability

We apply the classifiers associated with the annual false positive ratios ζ_{FP} of 0.1 and 0.05 per year on the test dataset $\mathcal{D}^{\mathrm{test,\,TP}}$ in order to obtain the probability of true detections for inclinometers and extensometers. Figure 5.5 illustrates the probability of true detections for the sensors where $\mathtt{I}_{\mathrm{L}}$ is the lateral inclinometer, $\mathtt{I}_1$ is the longitudinal inclinometer 1, $\mathtt{I}_3$ is the longitudinal inclinometer 3, and $\mathtt{E}$ is the extensometer. The three axes in each figure represent the covariates vector $\mathbf{x} = [|h_{\mathrm{a}}| \; w_{\mathrm{a}} \; \mathtt{N}_{t>t_{\mathrm{s}}}]^{\mathsf{T}}$ consisting in the absolute value of the anomaly magnitude $|h_{\mathrm{a}}|$, its duration w_{a}, and the number of time steps after the start of the anomaly $\mathtt{N}_{t>t_{\mathrm{s}}}$, which determines the delay in the anomaly detection. Figures 5.5a-5.5d show the probability of true detections for the annual false positive $\zeta_{\mathrm{FP}} = 0.05/\mathrm{year}$, while Figures 5.5e-5.5h are associated with the $\zeta_{\mathrm{FP}} = 0.1/\mathrm{year}$. The colors indicate the probability of true detections $\Pr(Y = +1 : \mathrm{TP}|\mathbf{x}) \in (0, 1)$.

Figure 5.6 shows the isosurfaces corresponding to a 90% probability of true positive detections extracted from the results obtained by TAGI and shown in Figure 5.5. The blue plane quantifies the minimum detectable anomaly magnitude $|h_a|_{\mathrm{min}}$. The yellow plane quantifies the minimum detectable anomaly growth rate per year $\psi_{\mathrm{min}} = \psi_0 \times |h_a|_{\mathrm{min}}$ as a function of its corresponding $|h_a|_{\mathrm{min}}$.

For a better illustration, we re-draw Figure 5.6 from the top view, i.e., $|h_a| - w_a$ plane perspective, and present the results in Figure 5.7. Figures 5.7a-5.7d correspond to the case for which the annual false positive ratio $\zeta_{\mathrm{FP}} = 0.05/\mathrm{year}$, while Figures 5.7e-5.7h are associated with $\zeta_{\mathrm{FP}} = 0.10/\mathrm{year}$. Analogously to the results obtained using reinforcement learning and presented in Section 4.3.3, the comparison between these two annual false positive ratio cases confirms that increasing the annual false positive ratio results in an increase in the true positive detections. Moreover, Figures 5.7e-5.7h which have higher annual false positive ratios, show earlier anomaly detections as opposed to Figures 5.7a-5.7d which has a lower ζ_{FP}. In addition, comparing the red areas for the two cases of ζ_{FP} reveals that the higher
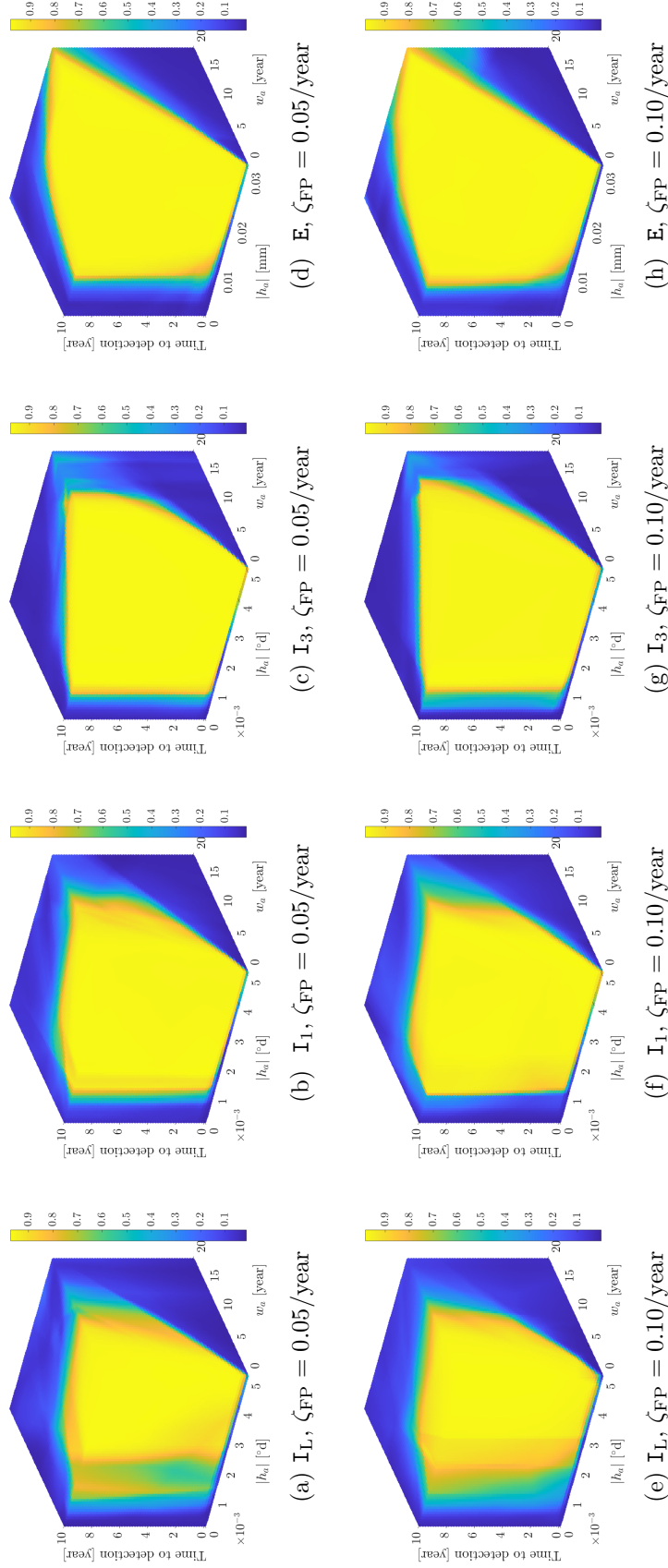
(a) I_L, $\zeta_{FP} = 0.05$/year

(b) I_1, $\zeta_{FP} = 0.05$/year

(c) I_3, $\zeta_{FP} = 0.05$/year

(d) E, $\zeta_{FP} = 0.05$/year

(e) I_L, $\zeta_{FP} = 0.10$/year

(f) I_1, $\zeta_{FP} = 0.10$/year

(g) I_3, $\zeta_{FP} = 0.10$/year

(h) E, $\zeta_{FP} = 0.10$/year

Figure 5.5 Illustration of the probability of the true positive detection for: (a-d) $\zeta_{FP} = 0.05$, and (e-h) $\zeta_{FP} = 0.1$/year. The sensor names are I_L: Lateral inclinometer, I_1: Longitudinal inclinometer 1, I_3: Longitudinal inclinometer 3, and E: Extensometer. In addition, the unit for the anomaly magnitude depends on the sensor type, that is mm for the extensometer and °d for the inclinometers.

(a) $\mathrm{I_L}$, $\zeta_{\mathrm{FP}} = 0.05$/year

(b) $\mathrm{I_1}$, $\zeta_{\mathrm{FP}} = 0.05$/year

(c) $\mathrm{I_3}$, $\zeta_{\mathrm{FP}} = 0.05$/year

(d) $\mathbf{E}$, $\zeta_{\mathrm{FP}} = 0.05$/year

(e) $\mathrm{I_L}$, $\zeta_{\mathrm{FP}} = 0.10$/year

(f) $\mathrm{I_1}$, $\zeta_{\mathrm{FP}} = 0.10$/year

(g) $\mathrm{I_3}$, $\zeta_{\mathrm{FP}} = 0.10$/year

(h) $\mathbf{E}$, $\zeta_{\mathrm{FP}} = 0.10$/year
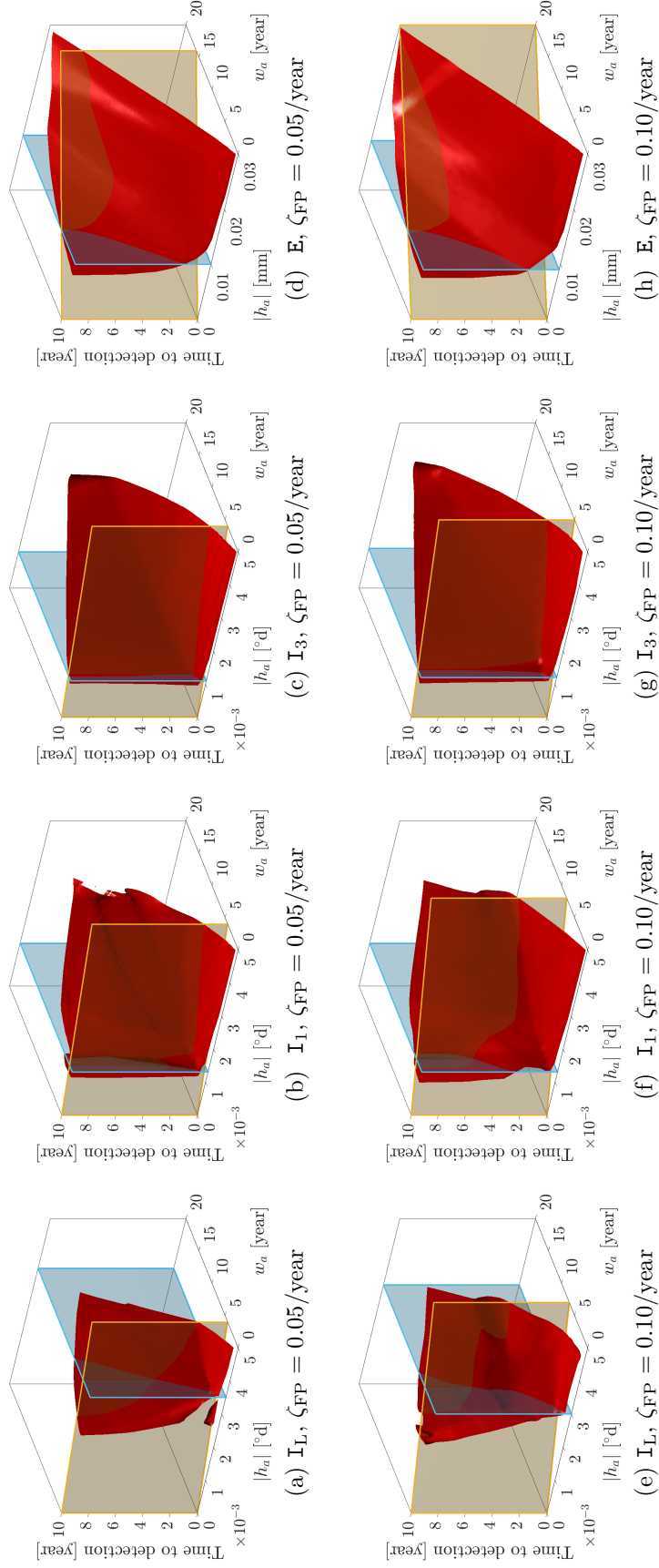
Figure 5.6 Illustration of the 90% probability for true positive detections for: (a-d) $\zeta_{\mathrm{FP}} = 0.05$, and (e-h) $\zeta_{\mathrm{FP}} = 0.1$/year. The sensor names are $\mathrm{I_L}$: Lateral inclinometer, $\mathrm{I_1}$: Longitudinal inclinometer 1, $\mathrm{I_3}$: Longitudinal inclinometer 3, and $\mathbf{E}$: Extensometer. The blue plane quantifies the minimum detectable anomaly magnitude $|h_a|_{\min}$. The yellow plane quantifies the minimum detectable anomaly growth rate per year $\psi_{\min} = \psi_0 \times |h_a|_{\min}$ as a function of its corresponding $|h_a|_{\min}$.
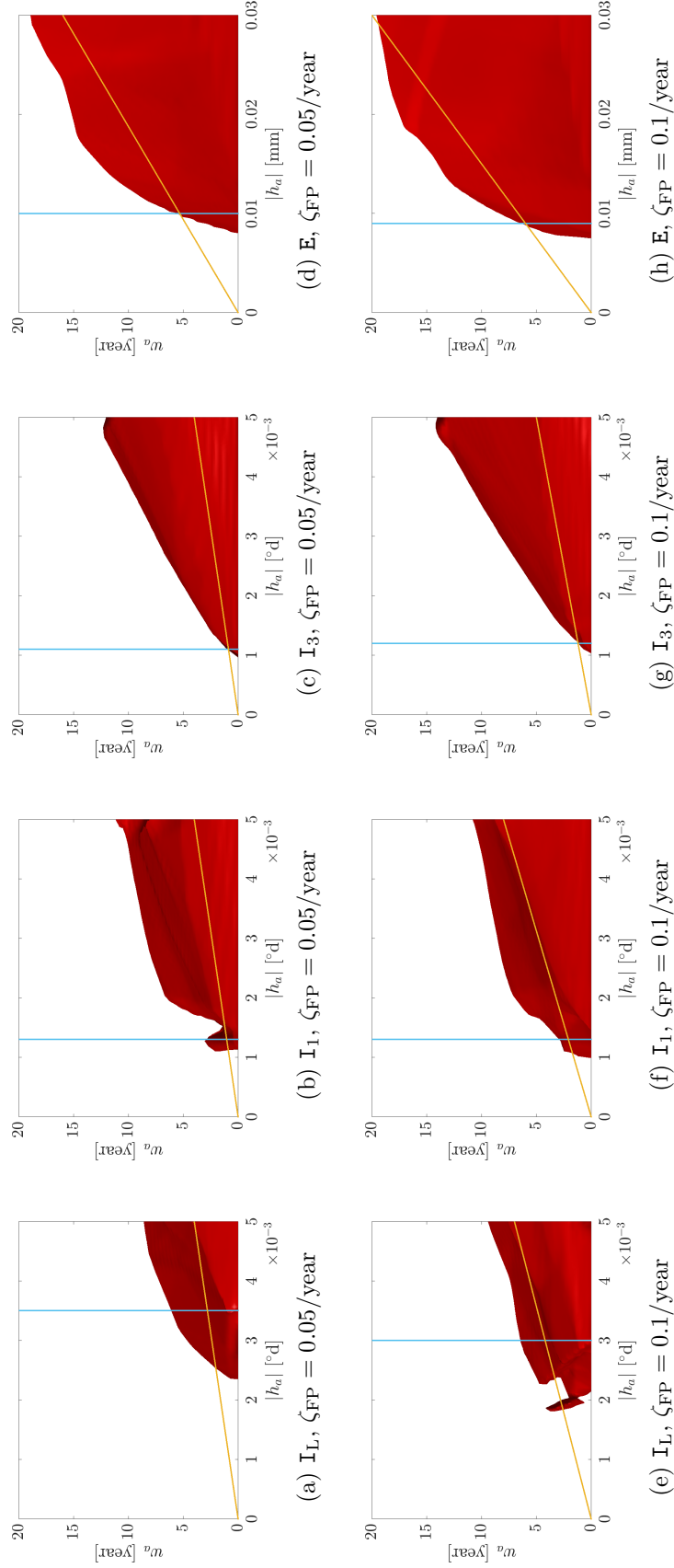
Figure 5.7 Comparison of the isosurface corresponding to 90% probability of true positive detections in $|h_a| - w_a$ plane for: (a–d) $\zeta_{\mathrm{FP}} = 0.05$, and (e–h) $\zeta_{\mathrm{FP}} = 0.1$/year. The sensor names are $\mathrm{I_L}$: Lateral inclinometer, $\mathrm{I_1}$: Longitudinal inclinometer 1, $\mathrm{I_3}$: Longitudinal inclinometer 3, and E: Extensometer. The blue line indicates the minimum detectable anomaly magnitude $|h_a|_{\mathrm{min}}$ with the probability of 90%. The red line indicates the minimum detectable anomaly growth rate ψ per year with the same probability of 90%.

annual false positive ratio results in an increase in the red areas with respect to the anomaly magnitude and duration. This implies that for the case with $\zeta_{\mathrm{FP}} = 0.10$/year, the agent is able to detect anomalies with a lower magnitudes and with longer duration.

Table 5.2 summarizes the minimum detectable anomaly magnitude and its growth rate. The first column shows the shorthand and complete name of the sensors. The second column is associated with the agents with the annual false positive ratios of $\zeta_{\mathrm{FP}} = 0.05$ and 0.1 per year. The third column identifies the minimum detectable anomaly magnitude $|h_a|_{\min}$, and the fourth column shows the constant ψ_0. The fifth column shows the minimum detectable anomaly growth rates $\psi_{\min}$ in years as a function of $|h_a|_{\min}$, i.e., $\psi_{\min} = |h_a|_{\min}$ per $\mathbb{N}_\psi$ years. For instance, for the agent corresponding to the lateral inclinometer with the annual false alarm ratio of $\zeta_{\mathrm{FP}} = 0.05$/year, the minimum anomaly magnitude must be equal or greater than $0.003°$d per $\mathbb{N}_\psi = 13$ years. In other words, an anomaly with a magnitude of $0.003°$d is detectable, if it develops in less than thirteen years.

Table 5.2 Minimum detectable anomaly magnitudes $|h_a|_{\min}$ and the the constant ψ_0 corresponding to the minimum detectable anomaly growth rates in year such that $\psi_{\min} = \psi_0 \times |h_a|_{\min}$. The fifth and last columns show the minimum detectable anomaly growth rates $\psi_{\min}$ in years for both imitation and reinforcement learning approaches as a function of $|h_a|_{\min}$, i.e., $\psi_{\min} = |h_a|_{\min}$ per $\mathbb{N}_\psi$ years.

Sensor ID and name	ζ_{FP}	Imitation learning			Reinforcement learning	
		$|h_a|_{\min}$	ψ_0	$\mathbb{N}_\psi$ [year]	$|h_a|_{\min}$	$\mathbb{N}_\psi$ [year]
$\mathtt{I_L}$: Lateral inclinometer	0.05	$0.0030\,°$d	0.08	13	$0.0017\,°$d	7
	0.10	$0.0035\,°$d	0.07	15	$0.0020\,°$d	8
$\mathtt{I_1}$: Longitudinal inclinometer 1	0.05	$0.0013\,°$d	0.19	6	$0.0013\,°$d	6
	0.10	$0.0013\,°$d	0.19	6	$0.0013\,°$d	6
$\mathtt{I_3}$: Longitudinal inclinometer 3	0.05	$0.0011\,°$d	0.23	5	$0.0010\,°$d	4
	0.10	$0.0012\,°$d	0.21	5	$0.0010\,°$d	4
$\mathtt{E}$: Extensometer	0.05	$0.0100\,$mm	0.15	7	$0.0100\,$mm	7
	0.10	$0.0090\,$mm	0.17	6	$0.0090\,$mm	6

In Table 5.2, the last two columns are associated with the minimum detectable anomaly magnitude and minimum detectable anomaly growth rates in years obtained from the reinforcement learning approach presented in Table 4.7. Comparing these two sets of values reveals that both approaches result in a similar anomaly detectability for all the sensors excepted for the lateral inclinometers. In this case, the reinforcement learning approach shows a better anomaly detectability with respect to the minimum detectable anomaly's magnitude $|h_a|_{\min}$. This is also confirmed by comparing Figures 5.7a and 5.7e associated with the imitation learning approach with the ones shown in Figures 4.19a and 4.19e corresponding to the reinforcement learning approach. Note that the comparison is limited out based on the pro-

posed simplified planes, while more complex functions could be used in order to compare the performance between two learning approaches. Moreover, comparing the maximum number of years for developing the minimum anomaly N_ψ between the RL and IL approaches reveals that the IL-based anomaly detection allows detecting anomalies that take a longer time to develop. In particular, this is the case for the anomaly detectability corresponding to the lateral inclinometer; The values for N_ψ are higher than the ones when the anomaly detection method is based on the reinforcement learning approach. However, this is at the expense of detecting anomalies with higher magnitude, where these magnitudes are approximately doubled when the anomaly detection method is based on the imitation learning approach.

5.4 Discussion

In the context of the anomaly detection problem, the goal of an agent in the imitation learning (IL) approach is to identify alarm triggering policies by imitating an expert's policy. IL-based anomaly detection suffers from few limitations; First, a demonstration set is provided by an expert that is typically a human for whom the proposed policy could be suboptimal by itself. As a result, the training of an agent is carried out with suboptimal decisions that will result in suboptimal solutions [94]. Here, the task of anomaly detection is relatively straightforward in the sense that an agent requires to make binary decision with respect to the presence of anomalies based on the expected value of the trend hidden state and the probability of the non-stationary regime, which form a two dimensional environment. Such a decision making is formulated according to Equation 5.2, and is applied on all the simulated episodes during the training of an agent. Therefore, we are confident that the policies issued by the expert are optimal.

In Section 2.2.5, we presented an example of how relying only on the probabilities of a non-stationary regime for triggering an alarm is prone to false or missed alarms. Here, we show such a limitation by quantifying probabilistically the anomaly detectability that would result from using a threshold on regime switching probabilities. To this end, we identify the probability threshold value $\pi_{\min}^{\mathbf{ns}, a^0}$ for the non-stationary regime probabilities $\pi^{\mathbf{ns}}$ that would act as the criterion for triggering an alarm. The superscript a^0 indicates that for the non-stationary probabilities $\pi^{\mathbf{ns}} > \pi_{\min}^{\mathbf{ns}, a^0}$ the action is to trigger an alarm. This threshold value is selected such that it results in an annual false positive ratio of $\zeta_{\mathrm{FP}} = 0.1/\mathrm{year}$, when it is applied to the false positive test datasets $\mathcal{D}^{\mathrm{test,\,FP}}$ presented in Table 4.4. In addition, the procedure for evaluating the probability of true positive detections using the true positive test datasets $\mathcal{D}^{\mathrm{test,\,TP}}$ is presented in Section 5.3.3.

Figure 5.8 illustrates the probability of true positive detections, i.e., $\Pr(Y = +1 : \mathrm{TP}|\mathbf{x}) \in$

$(0, 1)$, for the extensometers E. Figure 5.8a shows the probability of true detections corresponding to an annual false positive $\zeta_{\text{FP}} = 0.1/\text{year}$, where the threshold value is $\pi_{\min}^{\text{ns}, a^0} = 0.105$. The three axes represent the same covariates vector $\mathbf{x} = [|h_{\text{a}}|\ w_{\text{a}}\ \text{N}_{t>t_{\text{s}}}]^{\mathsf{T}}$ as presented in Section 5.3.3. Figure 5.8b shows the isosurface corresponding to a 90% probability of true positive detection extracted from Figure 5.8a.
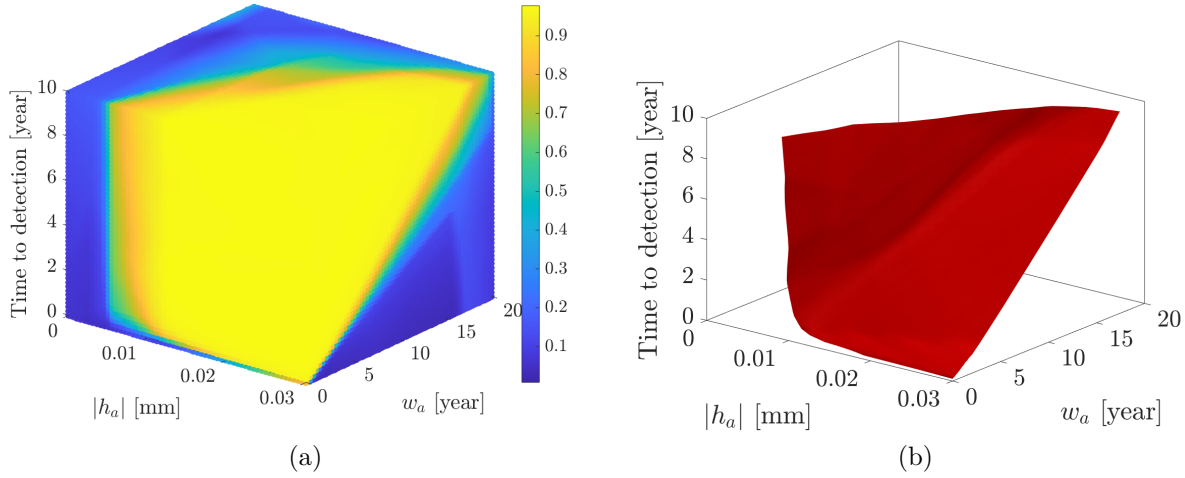
Figure 5.8 The anomaly detectability results using a threshold-based anomaly detection for the extensometers E: (a) probability of true positive detections with respect to prediction covariates vector $\mathbf{x}$, and (b) the isosurface corresponding to 90% probability of true positive detections. Figures correspond to an annual false positive $\zeta_{\text{FP}} = 0.1/\text{year}$, where the threshold value is $\pi_{\min}^{\text{ns}, a^0} = 0.105$.

We re-draw Figure 5.8b from the top view, i.e., $|h_a| - w_a$ plane perspective, and present it in Figure 5.9, along with the results that are obtained for the extensometer using reinforcement and imitation learning approaches which were presented in Figures 4.19h and 5.7h. Note that all the figures correspond to an annual false positive ratio of $\zeta_{\text{FP}} = 0.1/\text{year}$. The comparison between these three methods confirms that using the structural degradation's speed increases the anomaly detectability in the sense that reinforcement or imitation learning approaches enable detecting anomalies with lower magnitudes and longer durations, i.e., comparing the red areas in Figure 5.9.

| (a) Reinforcement learning | (b) Imitation learning | (c) Threshold-based |
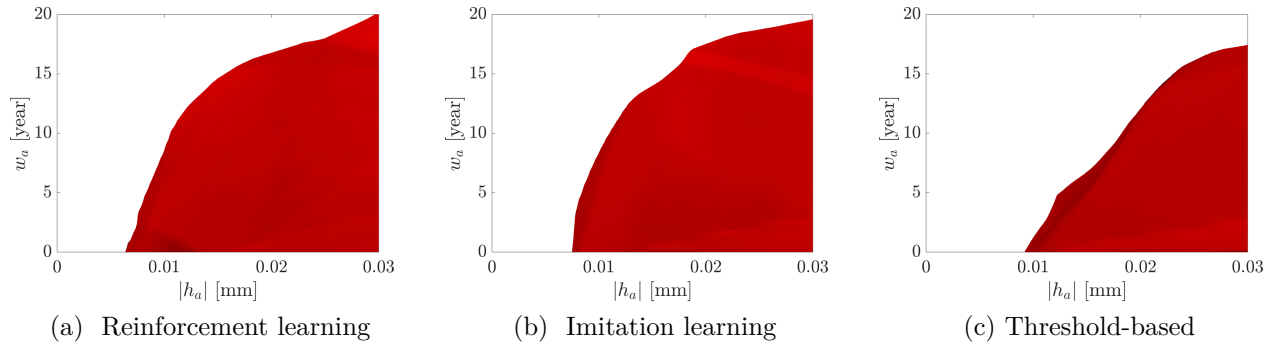
Figure 5.9 Comparison of the isosurfaces corresponding to 90% probability of true positive detections. Figures correspond to an annual false positive $\zeta_{\text{FP}} = 0.1/\text{year}$. The threshold value corresponding to Figure 5.9c is $\pi_{\min}^{\texttt{ns},\,a^0} = 0.105$.

5.5   Conclusion

This chapter addressed the problem of detecting anomalies as well as of quantifying the anomaly detectability using the imitation learning approach. The anomaly detection problem was formulated in the context of imitation learning for which an agent learns the alarm triggering policy using a set of demonstrations obtained from an expert's policy; This led to a supervised learning configuration where instead of allowing an agent to interact with an environment in order to collect the rewards, we built a classifier from the demonstrations, so an agent is able to imitate the expert's decisions. The demonstrations are obtained by employing the expert's decision with regard to the time to trigger an alarm on the same episodes simulated to train a reinforcement learning (RL) agent as described in Chapter 4.

We built a classifier for each sensor which resulted in a probabilistic classification between the action of triggering an alarm and not triggering an alarm. The minimum probability used to trigger an alarm for various annual false positive ratios was obtained using a grid-search procedure. We showed that the resulting decision boundaries are very close to each other indicating that the IL-based anomaly detection approach is not as robust as the RL-based approach on that aspect. Furthermore, we used the same methodologies developed in Chapter 4 in order to quantify the detectability of the anomalies for each sensor type, and present the corresponding two metrics, i.e., the annual false positive ratio and the probability of true positive detections; This chapter presented the resulting isosurfaces which identify the probability of true positive detections with respect to the anomaly magnitude and its duration, as well as the number of time steps an agent takes in order to detect an anomaly. It is confirmed that for a higher annual false positive ratio, a 90% probability of true positive detections is

achieved for an anomaly with lower magnitude and longer duration. Moreover, this chapter summarized the detectability of each sensor type with respect to two macroscopic quantities representing the minimum detectable anomaly magnitude and its growth rate. These two quantities give simple and interpretable information about the anomaly detectability of each sensor type corresponding to the current measurement system of the Jacques Cartier Bridge. It was found that the RL-based anomaly detectability yields similar results compared with the ones obtained from the IL-based one, except for the lateral inclinometer where the RL agent is able to detect anomalies with lower magnitudes. However, the computational cost for training an RL agent is higher than for training a classifier in the context of the IL, with respect to both number of episodes and the time to train the agents; We showed that the number of episodes required to train a classifier is an order of magnitude lower than the ones used to train RL agents. In addition, we showed that the total number of samples from the episodes can be reduced by orders of magnitudes, which accelerates training the classifier. Moreover, we presented a comparison between the anomaly detectability results obtained from the RL- and IL-based anomaly detection approaches and the ones based on relying on defining a threshold value for the probabilities of a non-stationary regime. It is confirmed that using the structural degradation's speed enables a better anomaly detection in the sense that we are able to detect anomalies with lower magnitudes and longer durations.

A more general challenge is to maximize the detectability of structural damages as well as the distinguishability among them. In this regard, a necessary step towards addressing this challenge is to first associate the magnitudes of anomalies to the severity levels of various structural damages, and second, to evaluate the performance of different sensor configurations in order to identify the one that maximizes the distinguishability among these damages. We showed that both anomaly detection methods based on the RL and IL approaches resulted in similar anomaly detectability for all sensors except for the lateral inclinometer. For this type of sensor RL-based anomaly detection showed better results in the sense that it enables detecting anomalies with lower magnitudes for both target annual false positive ratios. In Chapter 6 we will introduce additional methodologies in order to leverage minimum detectable anomaly magnitudes for detecting structural damages and distinguishing among these damages using different sensor configurations.

CHAPTER 6   Damage Detectability and Distinguishability

In Chapters 4 and 5, we have introduced methodologies in order to detect anomalies from structural responses, and to quantify the anomaly detectability with respect to the probability of true positive detections and the annual false positive ratio. When a damage occurs, structural responses depart from their normal state. Such a deviation is noticeable in the hidden state variables estimated from the measurements, and is detected by the reinforcement learning agents as described in Chapter 4 or the imitation learning ones described in Chapter 5. A more general challenge is to identify the cause of an anomaly's magnitude by associating them to the severity levels of structural damages. To this end, this chapter first presents a method to model structural damage scenarios and detect them using reinforcement learning-based agents described in Section 4.2.3. Furthermore, this chapter introduces a method in order to use the damage detectability of multiple sensors in order to measure the performance of a sensor configuration in distinguishing different scenarios. As a result, we are able to compare the performance of different sensor configurations in order to identify the one which maximizes the distinguishability between the damage scenarios.

6.1   Structural Damage Scenarios

We use the finite element (FE) method to build a model of a structure in order to study the changes in structural responses caused by different structural damage scenarios. The FE analysis quantifies the elements' deformations subject to various loading types. The deformations obtained without the application of damages are referred to as the *intact* state of the structure. Applying structural damages on a FE model results in new deformation values that are referred to as the *damaged* state of a structure. The corresponding nodal displacements and rotations associated with the deformations can be simulated at the locations where sensors are installed. The difference in nodal displacements and rotations between the intact and damaged states determines the change in structural responses due to each damage scenario. This change is associated with an anomaly's magnitude used in the process of defining the anomaly function's characteristics as explained in Section 4.1.1.

Figure 6.1a shows the deformation of an element due to a damage scenario. $\Delta\ell$ represents the change in structural responses before and after the deformation of an element for an extensometer, and $\Delta\theta$ indicates this change for an inclinometer. The difference between inclinations of the intact θ_{intact} and damaged θ_{damaged} state of the structure is $\Delta\theta = \theta_{\text{damaged}} - \theta_{\text{intact}}$. The extensometer readings are based on the relative displacements between the supports of
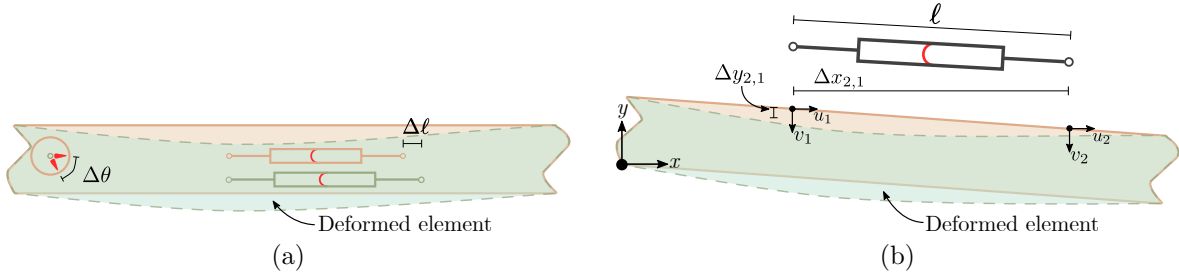
Figure 6.1 Illustration of (a) the deformation of an element and changes in the rotation of an inclinometer and the elongation of an extensometer, and (b) the displacements associated with the two nodes at the location of the supports of the extensometer on an structural element. $\Delta\theta$ and $\Delta\ell$ are the respective changes in structural responses at the location of an inclinometer and an extensometer. $\Delta x_{2,1} = x_2 - x_1$ and $\Delta y_{2,1} = y_2 - y_1$ represent the nodal coordinates differences in the respective x and y directions. u_i and vertical v_i are the respective nodal displacements at the location of the extensometer supports.

the extensometer. The length of the sensor is evaluated for the intact ℓ_{intact} and damaged ℓ_{damaged} states of the structure, and the difference is calculated as $\Delta\ell = \ell_{\text{damaged}} - \ell_{\text{intact}}$. Figure 6.1b shows the displacement for the two nodes corresponding to an extensometer supports. Each node has a coordinate and the difference between the nodal coordinates are $\Delta x_{2,1} = x_2 - x_1$ and $\Delta y_{2,1} = y_2 - y_1$. By defining the local horizontal u_i and vertical v_i displacements for each node $i \in \{1, 2\}$, the new length of the extensometer due to a deformation is obtained as $\ell = \sqrt{[\Delta x_{2,1} + (u_2 - u_1)]^2 + [\Delta y_{2,1} + (v_2 - v_1)]^2}$. The horizontal and vertical displacements are obtained from the FE analysis subject to different damage scenarios.

Table 6.1 shows a generic example of three different damage scenarios and the corresponding changes in the structural responses, i.e., $\Delta\theta$ and $\Delta\ell$, at the location of three sensors, namely s_1, s_2, and s_3. For each sensor, the minimum detectable damage scenario is obtained from

Table 6.1 Example of the changes in structural responses, i.e, $\Delta\theta$ and $\Delta\ell$, associated with s_1, s_2, and s_3 sensors due to three damage scenarios. The values in red are the ones detected via the comparison with $|h_a|_{\text{min}}$ obtained from corresponding agents.

Damage scenario	$\Delta\theta$		$\Delta\ell$
	s_1	s_2	s_3
#1	+0.004	+0.005	−0.005
#2	+0.003	+0.004	−0.020
#3	−0.003	−0.006	+0.010
$|h_a|_{\text{min}}$	0.003	0.003	0.01

$|h_a|_{\text{min}}$ values defined in Section 4.2.3. For instance, given a $|h_a|_{\text{min}} = 0.01$ for the exten-

someter s_3, damage scenarios shown in red in the fourth column in Table 6.1 are deemed to be detectable. In addition, considering that the minimum detectable anomaly magnitude for inclinometers is $|h_a|_{\min} = 0.003$, the sensors s_1 and s_2 would be able to detect all damage scenarios.

6.2 Cosine Similarity

In Section 6.1, we presented a method to generate damage scenarios using physics-based models and compare the induced changes in structural responses with the minimum detectable anomaly magnitude $|h_a|_{\min}$ obtained from a RL agent. This procedure allows to draw conclusions on whether or not these changes are detectable when relying on each sensor separately. A more general challenge is to choose a sensor configuration that maximizes the damage detectability and that can best distinguish between damage scenarios. To this end, we use the *cosine similarity* [95] to quantify the distinguishability between damage scenarios; It measures the similarity between two vectors in a multidimensional space as the cosine of the angle between them; If the two vectors are in the same direction, the cosine of the angle results in the maximum similarity with a value of $+1$, while if they are in opposite directions the similarity value is minimum at -1. Therefore, the objective is to choose a sensor configuration resulting in minimum cosine similarities between different damage scenarios in order to maximize the distinguishability between them.

In order to illustrate this point, consider the three sensors s_1, s_2, and s_3 presented in Table 6.1. There are seven possible sensor configurations, i.e., $\{s_1\}$, $\{s_2\}$, $\{s_3\}$, $\{s_1, s_2\}$, $\{s_1, s_3\}$, $\{s_2, s_3\}$, and $\{s_1, s_2, s_3\}$. Each configuration results in a vector that represents the changes in structural responses observed by the sensors due to a damage scenario. For instance, if all three sensors in Table 6.1 are chosen, each row in the table corresponds to a damage scenario vector in a three dimensional space, where each dimension corresponds to the structural response at the location of a sensor. These vectors are

$$\mathbf{d}^1 = [+0.004 \ + 0.005 \ 0]^\intercal,$$
$$\mathbf{d}^2 = [+0.003 \ + 0.004 \ - 0.020]^\intercal,$$
$$\mathbf{d}^3 = [-0.003 \ - 0.006 \ + 0.010]^\intercal,$$

where the superscripts identify the damage scenarios. Note that the third element of the vector $\mathbf{d}^1$ is zero because this damage scenario is not detectable by the RL agent at the location of the sensor s_3. If the cosine similarity between any pair of these vectors is close to one, the corresponding damage scenario is least distinguishable, regardless of the minimum detectable

anomaly magnitude obtained from RL agents. On the other hand, if the cosine similarity is close to -1, it is an indication that the two damage scenarios are highly distinguishable for that sensor configuration. Therefore, there is a need to compare these similarity values while accounting for the anomaly detectability in order to identify the best sensor configuration.

Mathematical Formulation

For a subset $\varrho \subseteq \{s_1, s_2, \ldots, s_{\mathtt{N}_{\text{sensor}}}\}$ from a set of $\mathtt{N}_{\text{sensor}}$ sensors, a *damage vector* $\mathbf{d}^i$ from a set of damage scenarios $\varphi = \{\mathbf{d}^1, \mathbf{d}^2, \ldots \mathbf{d}^{\mathtt{Q}}\}$ is defined as

$$\mathbf{d}^i = \begin{bmatrix} d_1^i & d_2^i & \ldots & d_{\mathtt{S}}^i \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{\mathtt{S} \times 1}, \; \mathtt{S} \leq \mathtt{N}_{\text{sensor}}, \tag{6.1}$$

where d_k^i, $k \in \{1, 2, \ldots \mathtt{S}\}$ is a value representing the change in structural responses obtained at the location of a sensor s_k, by applying the damage scenario i on a model of the structure. Note that an element of a damage vector $d_k^i = 0$ if the damage i is not detectable by the corresponding RL agent. Before evaluating cosine similarities, we modify the damage scenario vector such that

$$\hat{\mathbf{d}}^i = \text{sign}(\mathbf{d}^i) \odot \lfloor |\mathbf{d}^i \odot \mathbf{h}| \rfloor, \tag{6.2}$$

where, $\text{sign}(.)$ extracts the sign of a vector, $\odot$ is the element-wise vector multiplication, $\lfloor . \rfloor$ is the floor operator returning largest integer less than or equal to each element of a vector, and $|.|$ is the absolute value of a vector. In Equation 6.2, $\mathbf{h}$ is a vector consisting in the inverse of the minimum detectable anomaly magnitude obtained for each sensor. It is obtained from the anomaly detection analysis described in Section 4.2.3 following

$$\mathbf{h} = \begin{bmatrix} \dfrac{1}{|h_a|_{\text{min},1}} & \dfrac{1}{|h_a|_{\text{min},2}} & \cdots & \dfrac{1}{|h_a|_{\text{min},\mathtt{S}}} \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{\mathtt{S} \times 1}. \tag{6.3}$$

As a result, the modified damage scenario vectors $\hat{\mathbf{d}}^i$ consists only in integer values such that: (i) each element of the vector is a damage scenario represented as an integer multiplier of the corresponding $|h_a|_{\text{min}}$, and (ii) for the cases that a damage scenario value $|d_k^i| < |h_a|_{\text{min},k}$, Equation 6.2 returns a zero value indicating the damage scenario i is not detectable by the sensor k. For the example presented in Table 6.1, the modified damage scenario vectors are

$$\hat{\mathbf{d}}^1 = [+1 \; +1 \; 0]^{\mathsf{T}},$$
$$\hat{\mathbf{d}}^2 = [+1 \; +1 \; -2]^{\mathsf{T}},$$
$$\hat{\mathbf{d}}^3 = [-1 \; -2 \; +1]^{\mathsf{T}}.$$

The *cosine similarity matrix* $\mathbf{P} \in \mathbb{R}^{\mathbb{Q} \times \mathbb{Q}}$ is calculated for all modified damage scenario vector pairs following

$$[\mathbf{P}]_{ij} = \frac{\hat{\mathbf{d}}^i \cdot \hat{\mathbf{d}}^j}{\|\hat{\mathbf{d}}^i\| \|\hat{\mathbf{d}}^j\|} \in (-1, +1), \, \forall i, j \in \{1, 2, \ldots, \mathbb{Q}\}, \tag{6.4}$$

where, $\cdot$ is the dot product and $\|.\|$ is the norm of a vector, i.e., $\|\hat{\mathbf{d}}^i\| = \sqrt{\sum_{k=1}^{\mathbb{S}} \left[\hat{\mathbf{d}}^i\right]_k^2}$.

We define a metric in order to evaluate the distinguishability between a set of damages that is associated with a measurement system configuration. This metric is the average of the cosine similarity matrix values $[\mathbf{P}]_{ij}$ excluding the diagonal elements, which is defined as

$$\bar{p} = \sum_i \sum_j [\mathbf{P}]_{ij} / (\mathbb{Q}^2 - \mathbb{Q}), \, i \neq j. \tag{6.5}$$

In this setup, the best sensor configuration with respect to the damage distinguishability corresponds to the configuration ϱ^* that minimizes $\bar{p}$. In other words, we evaluate the cosine similarity matrix for all possible combinations of sensors, and compare the resulting $\bar{p}$ in order to select the configuration which can detect all of the damage scenarios and which has the minimum $\bar{p}$. Therefore, the optimal sensor configuration with respect to the cosine similarity matrix can be summarized by

$$\varrho^* = \underset{\varrho}{\arg\min} \, \bar{p}. \tag{6.6}$$

In addition, in this thesis, we discard the cosine similarities corresponding to the same type of the damage scenario pairs with different severities.

6.2.1   Cosine Similarity Application

Figure 6.2 illustrates the cosine similarity matrices associated with Table 6.1 for different sensor configurations. The matrices are symmetrical, so only their lower triangle part are taken into account for evaluating $\bar{p}$. The symbol $\emptyset$ in Figure 6.2b indicates that the damage scenario #1 is not detectable, i.e., $|d_k^i| < |h_a|_{\min,k}$, hence the corresponding damage scenario pairs are discarded. According to Equation 6.6, Figure 6.2e corresponds to the configuration $\{s_2, s_3\}$, which results in the minimum $\bar{p} = -0.42$, while being able to detect all damage scenarios.

Figure 6.3 illustrates the same three modified damage scenario vectors corresponding to Figure 6.2 in two and three dimensional spaces. In Figure 6.3a, only choosing the sensors s_1 and s_2 results in a two-dimensional space where $\hat{\mathbf{d}}^1$ and $\hat{\mathbf{d}}^2$ are in the same direction.
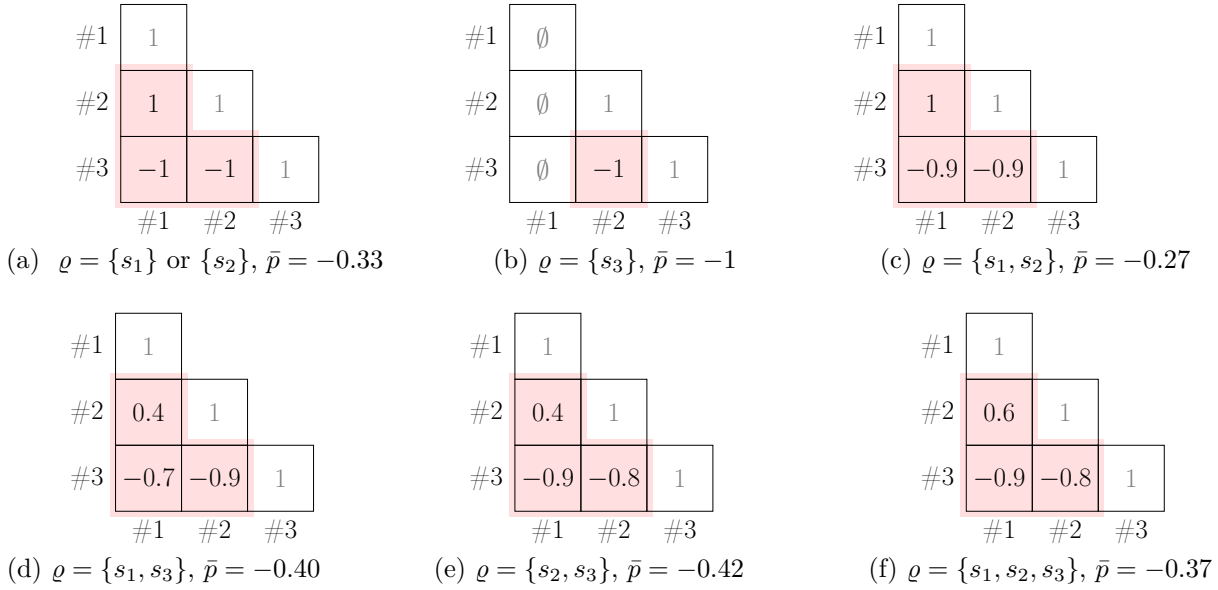
Figure 6.2 Illustration of the cosine similarity matrices $\mathbf{P}$ for different sensor configuration ϱ. Each configuration is a subset from $\{s_1, s_2, s_3\}$ a set consisting in three sensors. There are three damage scenarios numbered from 1 to 3. Only the lower triangle part of the matrices are taken into account for evaluating the average similarity metric $\bar{p}$. The optimal sensor configuration is correspond to the one with the minimum average similarity of $\bar{p} = -0.42$.

Therefore, the cosine similarity value is maximum, which indicates that we cannot distinguish between these two damage scenarios. In contrast, the damage scenario vector $\hat{\mathbf{d}}^3$ is in the opposite direction from $\hat{\mathbf{d}}^1$ and $\hat{\mathbf{d}}^2$ resulting in a negative cosine similarity. It indicates that the damage scenario #3 is distinguishable from #1 and #2. This is confirmed with the numerical values obtained in Figure 6.2c; The cosine similarity between the damage scenarios #1 and #2 is $+1$, while it is -0.9 between #1 and #3 as well as between #2 and #3. Using sensor s_3 in the sensor configurations shown in Figures 6.3b and 6.3c relaxes the limitation regarding the distinguishability between the damage scenarios #1 and #2. This is due to the fact that $\hat{\mathbf{d}}^1$ and $\hat{\mathbf{d}}^2$ vectors do not share the same value in the s_3 dimension. However, this is not the case for damage scenario #3; The comparison between Figure 6.3b and Figure 6.3c reveals that the distinguishability between $\hat{\mathbf{d}}^1$ and $\hat{\mathbf{d}}^3$ increases, while it reduces between $\hat{\mathbf{d}}^2$ and $\hat{\mathbf{d}}^3$. This is confirmed numerically in Figure 6.2d and 6.2e, for which the cosine similarity changes from -0.7 to -0.9 for $\hat{\mathbf{d}}^1$ and $\hat{\mathbf{d}}^3$, while it changes from -0.9 to -0.8 for $\hat{\mathbf{d}}^2$ and $\hat{\mathbf{d}}^3$. Figure 6.3d presents the configuration where all three sensors are used. The corresponding numerical values are presented in Figure 6.2f. The comparison between using three and two sensors shows that using more sensors can alter the distinguishability between a pair of damage scenarios; The comparison between Figure 6.2c and Figure 6.2f shows that
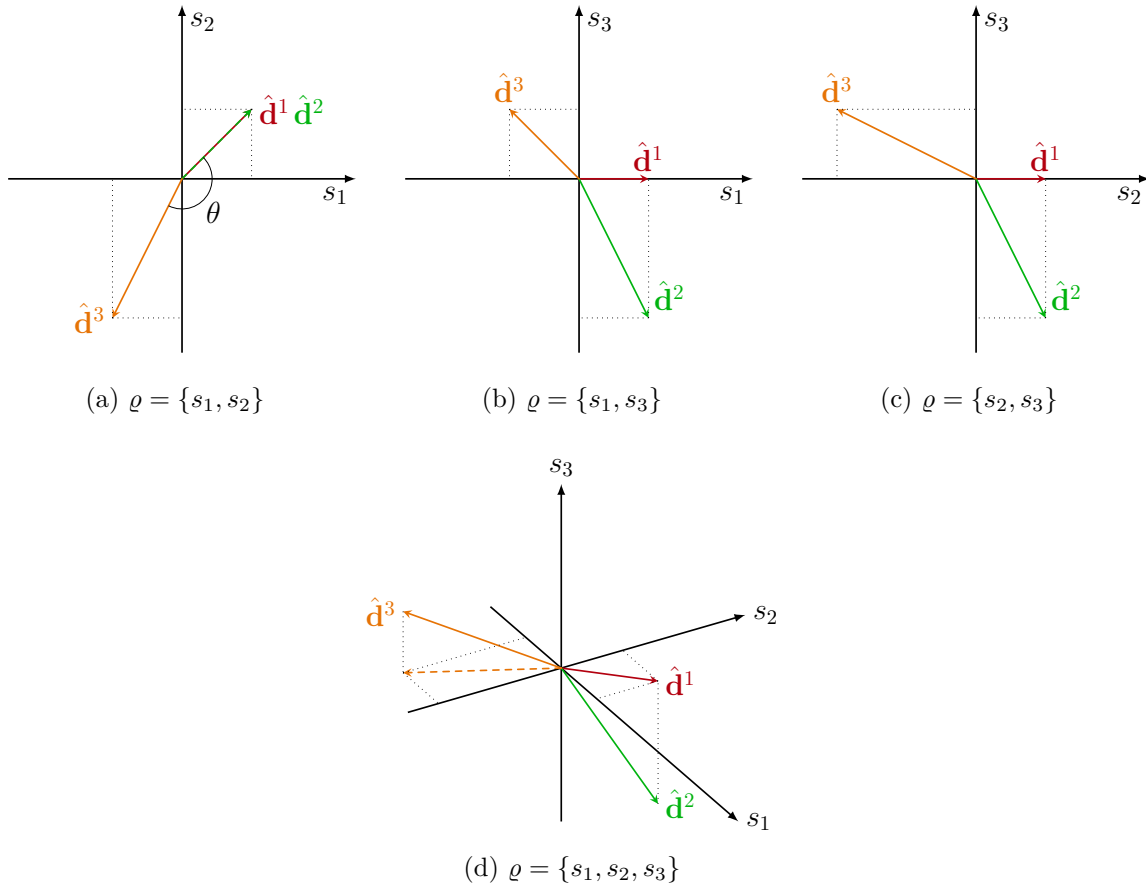
Figure 6.3 An example of the modified damage vectors $\hat{\mathbf{d}}^1$, $\hat{\mathbf{d}}^2$, and $\hat{\mathbf{d}}^3$ in: (a) two dimensional space $\varrho = \{s_1, s_2\}$, (b) two dimensional space $\varrho = \{s_1, s_3\}$, (c) two dimensional space $\varrho = \{s_2, s_3\}$, and (d) three dimensional space $\varrho = \{s_1, s_2, s_3\}$. The cosine similarity between each pair of vectors can be calculated depending on the sensor configuration. In some cases the vectors are in the same direction resulting in the maximum cosine similarity value of $+1$, while in some case it yields the minimum value of -1.

adding a sensor improves the distinguishability between the damage scenarios #1 and #2, i.e., the cosine similarity changes from $+1$ to $+0.6$. On the other hand, the comparison between Figure 6.2e and Figure 6.2f indicates that the distinguishability between #1 and #2 is reduced as the cosine similarity changes from $+0.4$ to $+0.6$. Using the metric $\bar{p}$ enables maximizing the distinguishability between different damage scenario pairs. In the next section we present the results associated with various damage scenarios and the distinguishability of different sensor configurations on the Jacques Cartier Bridge.

6.3 Results

In this section, first we present the structural damage scenarios modeled for the Jacques Cartier Bridge in order to study their effects on the three selected spans. Afterwards, we use the minimum detectable anomaly magnitude presented in Chapter 4 in order to measure the capacity of each sensor at detecting structural damages. Finally, we identify the sensor configuration that maximizes the distinguishability between structural damage scenarios.

6.3.1 Jacques Cartier Bridge Damage Simulation

As explained in Chapter 3, all three spans of the Jacques Cartier Bridge are instrumented with the same sensor configuration. Figure 6.4 illustrates the FE model that was created for simulating damage scenarios for a span of the bridge. The extensometer and three inclinometers are located on the downstream side of the bridge coinciding with the location of the sensors on the actual span. Bearing devices are located on both side of the span as depicted in Figure 6.4.
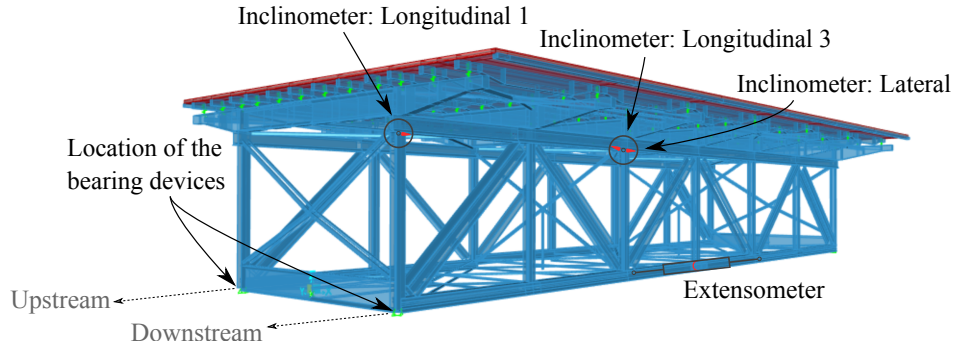
Figure 6.4 Illustration of the FE model for the spans and the locations of the sensors and bearing devices.

In this research, we consider that the structural responses for the intact state of the structure is due to static loading induced from the elements' self-weight. This is because the quantities we are interested in are the changes in the structural responses between the intact and damaged state of a structure, i.e., $\Delta\theta$ and $\Delta\ell$.

We consider three categories of structural damage scenarios that are affecting each span globally. They are *corrosion* (COR), *partial settlement* (PST), and *bearing device lockage* (BDL). The corrosion of a span is modeled as a reduction in the cross-section area of steel elements. Partial settlements are modeled as relative vertical displacements of the supports.

Partial settlements are considered in two setups namely, *uniform* (U) and *non-uniform* (NU). The uniform settlement corresponds to the settlement of one support relative to the others. The side for which the vertical settlements are applied is the one where the bearing devices are located as shown in Figure 6.4. The non-uniform settlements are associated with the relative settlement of only one support that is located on the upstream side of the bridge at the location of the bearing device. For simulating a bearing device lockage, we use a spring at the location of the supports in order to connect the fixed horizontal boundary condition and the bridge members. The stiffness of the spring is bounded between $k_{\min} = 0$ and $k_{\max}$. They represent the situation where they can behave like a roller or a fixed support, respectively. When the loading is applied on the span, the corresponding horizontal displacements of the bearing devices change depending on the value of the spring's stiffness. In order to model various scenarios, different values of stiffness are selected and the reduction in the average of the bearing device horizontal displacement is evaluated in percentage. Bearing device lockage is modeled in two setups namely, *uniform* and *non-uniform.* The uniform configuration corresponds to applying a scenario on both bearing devices. The non-uniform configuration is associated with applying a damage scenario only on the bearing device located on the upstream side of the bridge.

Figure 6.5 illustrates an example of the effect caused by a bearing device lockage damage. Figure 6.5a shows the intact state of the span for which the deformations are due to the elements' self-weight. The sensor names are I_L: Lateral inclinometer, I_1: Longitudinal inclinometer 1, I_3: Longitudinal inclinometer 3, and E: Extensometer. In addition, the numbers shown beside each sensor name are the value of either $\Delta\theta$ or $\Delta\ell$, where the units depend on the sensor type, that is °d for the inclinometers and mm for the extensometer. Figure 6.5b shows the damaged state of the span for which a 36.3% lockage is applied on the bearing device located on the upstream side the span. Accordingly, the corresponding relative rotations and elongation, i.e., $\Delta\theta$ and $\Delta\ell$, at the location of the sensors are changed. As it will be explained in the next section, these relative changes in the structural responses are the basis for identifying the damage scenarios that are detectable by each sensor.

Figure 6.6 illustrates the effects of the damage scenarios on the changes in structural responses at the location of the inclinometers and the extensometer. The corresponding numerical values are provided in Appendix C. Despite the interventions associated with the steel maintenance program described in Section 3.2, our analysis of the damage scenarios has revealed that the changes in structural responses for the three spans are indistinguishable. In Figure 6.6, the uniform damage scenarios are indicated by the red color, while the blue ones correspond to the non-uniform damage scenarios. Figures 6.6b and 6.6k indicate that the uniform partial settlement damage scenarios do not have any effect on the struc-

(a) Intact structure   (b) Damaged structure
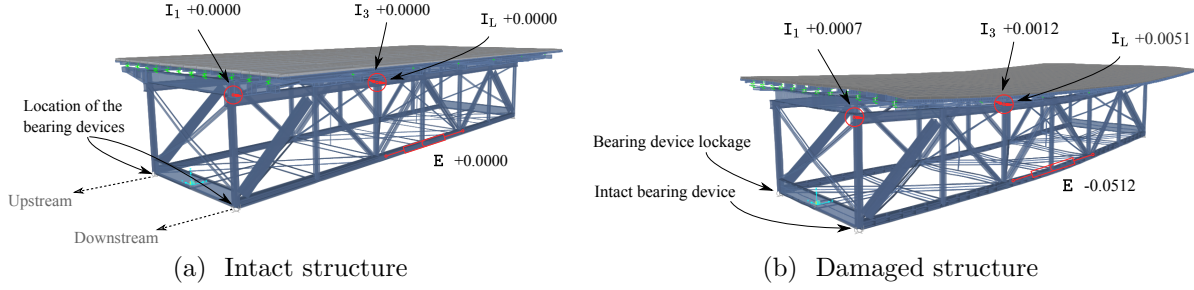
Figure 6.5 Effects of a 36.3% non-uniform bearing device lockage on the changes in relative structural responses $\Delta\theta$ and $\Delta\ell$ at the location of the inclinometers and the extensometer. The numerical values are provided in Appendix C. Also, $\mathtt{I_L}$: Lateral inclinometer, $\mathtt{I_1}$: Longitudinal inclinometer 1, $\mathtt{I_3}$: Longitudinal inclinometer 3, and $\mathtt{E}$: Extensometer.

tural responses at the location of the lateral inclinometer and extensometer. In addition, Figure 6.6c shows that for the lateral inclinometer, the effect of a bearing device lockage in the uniform setup is negligible compared with the non-uniform one. However, this sensor is the most sensitive inclinometer towards a non-uniform bearing device lockage in the sense that it yields the maximum rotation for any damage scenario of this type. As it is seen in Figures 6.6d and 6.6f, the sensor $\mathtt{I_1}$ is the most sensitive sensor among inclinometers for detecting corrosion and uniform bearing device lockage damage scenarios. The sensors $\mathtt{I_1}$ and $\mathtt{I_3}$ show a similar damage detectability with respect to both uniform and non-uniform partial settlements damage scenarios as depicted in Figures 6.6e and 6.6h.

The results shown in Figure 6.6 display the effects of changes in structural responses for each sensor based on the FE model. However, in order to identify the best sensor configuration, two factors must be accounted for: (i) the damage detectability of each sensor according to $|h_a|_{\min}$, and (ii) the distinguishability between different damage scenarios. The former is achieved by comparing $|h_a|_{\min}$ with the results from Figure 6.6, and the latter is addressed using the cosine similarity metric described in Section 6.2.

6.3.2   Jacques Cartier Bridge Damage Detectability

In order to identify damages that are detectable by the RL-based agents, we compare the minimum detectable anomaly magnitude $|h_a|_{\min}$ obtained from the Table 4.7 with the values shown in Figure 6.6. Table 6.2 shows the results of their comparison for all sensors, where the red color indicates the damage scenarios that are detectable by the agents with an annual false alarm ratio ζ_{FP} of 0.1/year. The second column introduces five *damage types* used in order to analyze the changes in structural responses. The third column presents the abbreviation for damage scenarios names. For instance, $\mathtt{BDL-NU-11\%}$ is the damage scenario associated

with 11% bearing device lockage in the non-uniform setup. The fourth to seventh columns show the changes in structural responses at the location of the sensors due to each damage
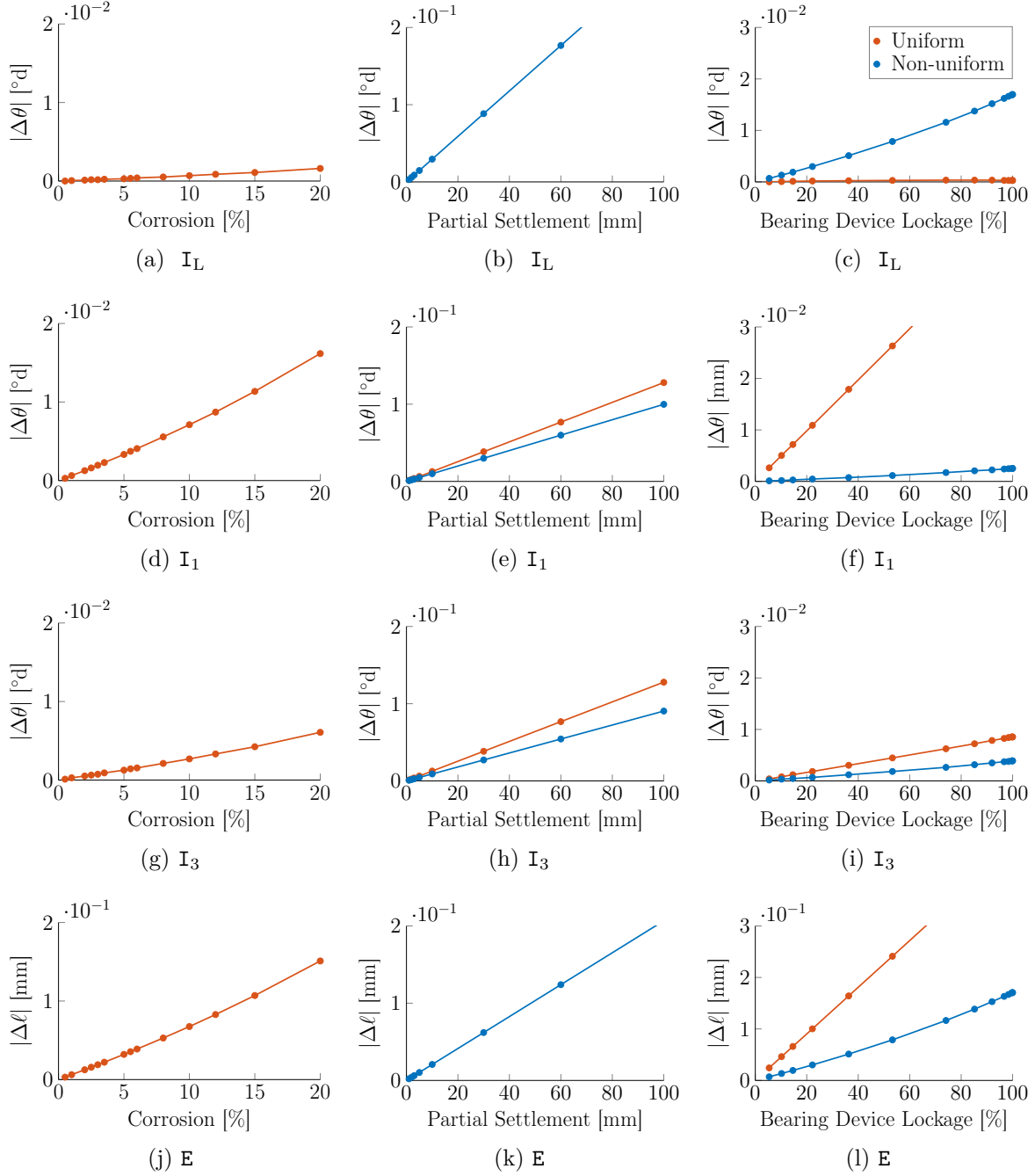
Figure 6.6 Effects of the damage scenarios on the changes in structural responses $\Delta\theta$ and $\Delta\ell$ at the location of the inclinometers and the extensometer. The numerical values are provided in Appendix C. Also, $\texttt{I}_\text{L}$: Lateral inclinometer, $\texttt{I}_1$: Longitudinal inclinometer 1, $\texttt{I}_3$: Longitudinal inclinometer 3, and $\texttt{E}$: Extensometer

scenario. For brevity, we use simplified names for the sensors, that is I_L for the lateral inclinometer, I_1 for the longitudinal inclinometer 1, I_3 for the longitudinal inclinometer 3, and E for the extensometer.

In Table 6.2, the severity of each damage type has a lower and an upper bound. For example, these severities are 2% and 5% for the corrosion. The lower bound severity is associated with the damage scenario that is detectable by at least one sensor; Severities smaller than the lower bound are not detectable by any sensor. On the other hand, the upper bound severity corresponds to the damage scenario for which the number and types of sensors that are able to detect severities larger than the upper bound remain the same. In other words, the upper bounds indicates the smallest severity level that is detectable by all the sensors. For example, the minimum severity corresponding to the corrosion is detectable by the inclinometer: Long-1 and extensometer; None of the sensors are able to detect the damages associated with a corrosion of less than 2%. A corrosion of 5% is the upper bound severity. It is detectable by the inclinometers Long-1 and Long-3 as well as the extensometer meaning that a corrosion larger than 5% is detectable with the same sensors. Note that the lateral inclinometer cannot detect this type of damage.

Table 6.2 Changes in structural responses, i.e, $\Delta\theta$ and $\Delta\ell$ due to different damage scenarios, associated with sensors installed on the JCB. The damage scenario names consist in two or three parts separated by dashes. They represent the damage's type, setup, and severity such that COR: corrosion, BDL: bearing device lockage, PST: partial settlement, U: uniform, and NU: non-uniform. Also, I_L: Lateral inclinometer, I_1: Longitudinal inclinometer 1, I_3: Longitudinal inclinometer 3, and E: Extensometer.

	Damage Scenario		$\Delta\theta$, °d			$\Delta\ell$, mm
ID	Type	Abbreviation	I_L	I_1	I_3	E
1	Corrosion	$\mathtt{COR}-2\%$	+0.0001	+0.0013	+0.0005	+0.0125
2		$\mathtt{COR}-5\%$	+0.0003	+0.0033	+0.0013	+0.0321
3	Uniform bearing	$\mathtt{BDL}-\mathtt{U}-5.4\%$	0.0000	−0.0026	−0.0004	−0.0244
4	device lockage	$\mathtt{BDL}-\mathtt{U}-14.6\%$	+0.0001	−0.0072	−0.0012	−0.0660
5	Non-uniform	$\mathtt{BDL}-\mathtt{NU}-10.2\%$	+0.0013	+0.0002	+0.0003	−0.0135
6	bearing device	$\mathtt{BDL}-\mathtt{NU}-22.2\%$	+0.0030	+0.0005	+0.0007	−0.0301
7	lockage	$\mathtt{BDL}-\mathtt{NU}-36.3\%$	+0.0051	+0.0007	+0.0012	−0.0512
8		$\mathtt{BDL}-\mathtt{NU}-74.1\%$	+0.0116	+0.0017	+0.0026	−0.1164
9	Uniform partial settlement	$\mathtt{PST}-\mathtt{U}-1\,\mathrm{mm}$	0.0000	−0.0013	−0.0013	0.0000
10	Non-uniform	$\mathtt{PST}-\mathtt{NU}-1\,\mathrm{mm}$	−0.0030	−0.0010	−0.0009	+0.0021
11	partial settlement	$\mathtt{PST}-\mathtt{NU}-2\,\mathrm{mm}$	−0.0059	−0.0020	−0.0018	+0.0041
12		$\mathtt{PST}-\mathtt{NU}-5\,\mathrm{mm}$	−0.0147	−0.0050	−0.0045	+0.0103
		$|h_a|_{\min}$	0.0020	0.0013	0.0010	0.0090

From Table 6.2, one can compare the damage detectability of each sensor individually. For instance, the best sensor is the extensometer as it detects more damage scenarios than any other. If we want to be able to detect all 12 damage scenarios, we would need to use the extensometer, the lateral inclinometer, and either the longitudinal inclinometer 1 or 3. In order to illustrate this point, consider the two scenarios $\mathtt{COR} - 2\%$ and $\mathtt{PST} - \mathtt{NU} - 5\,\mathrm{mm}$ for the extensometer shown in the first and last row of Table 6.2. If the corresponding agent triggers an alarm for which the anomaly magnitude is $|h_a| = 0.02$, it is not possible to identify the cause of this anomaly only from the extensometer. Nevertheless, we can rely on multiple sensors in order to distinguish between these two scenarios. For example, by using the lateral inclinometer, one can realize that the $\mathtt{PST} - \mathtt{NU} - 5\,\mathrm{mm}$ scenario is the cause of the anomaly because the lateral sensor cannot detect the $\mathtt{COR} - 2\%$ scenario. In the next section, we present a systematic analysis in order to obtain an optimal sensor configuration that maximizes the distinguishability between the damage scenarios for the Jacques Cartier Bridge using the cosine similarity metric.

6.3.3 Jacques Cartier Bridge Measurement System Design

We follow the procedure described in Section 6.2 in order to evaluate cosine similarity of all the sensor configurations for the 12 damage scenarios presented in Table 6.2. Figure 6.7 shows the cosine similarity matrices where the color assigned for each cell in a heat-map represents a cosine similarity value between two damage scenarios that varies from -1 to $+1$. The white cells represent the null values that are related to situations where a damage scenario is not detected by the sensor configuration. In such cases, the entire corresponding row and column are null. Figures 6.7a-6.7l are the cases containing null values that are not considered in our analysis. In addition, we do not consider the cosine similarity values that either belong to the diagonal of the matrix or are associated with damage scenario pairs that are of the same type but with different severities. We discarded these cases because once a damage is detected by a sensor, its severity is always identified by the sensor, while the usage of the cosine similarity is only to evaluate the capacity to distinguish between damage scenarios. Among all the heat-maps shown in Figure 6.7, only Figures 6.7m-6.7o correspond to the sensor configurations that are able to detect all 12 damage scenarios. Therefore, the evaluation of the average cosine similarity $\bar{p}$ is limited to these three figures, that is -0.12 for $\varrho = \{\mathtt{I_L}, \mathtt{I_1}, \mathtt{E}\}$, -0.13 for $\varrho = \{\mathtt{I_L}, \mathtt{I_3}, \mathtt{E}\}$, and -0.12 for $\varrho = \{\mathtt{I_L}, \mathtt{I_1}, \mathtt{I_3}, , \mathtt{E}\}$.

According to Equation 6.6, the best measurement system configuration corresponds to Figure 6.7n resulting in the minimum average cosine similarity of $\bar{p} = -0.13$. This configuration consists in the lateral inclinometer, the longitudinal inclinometer 3, and the extensometer,
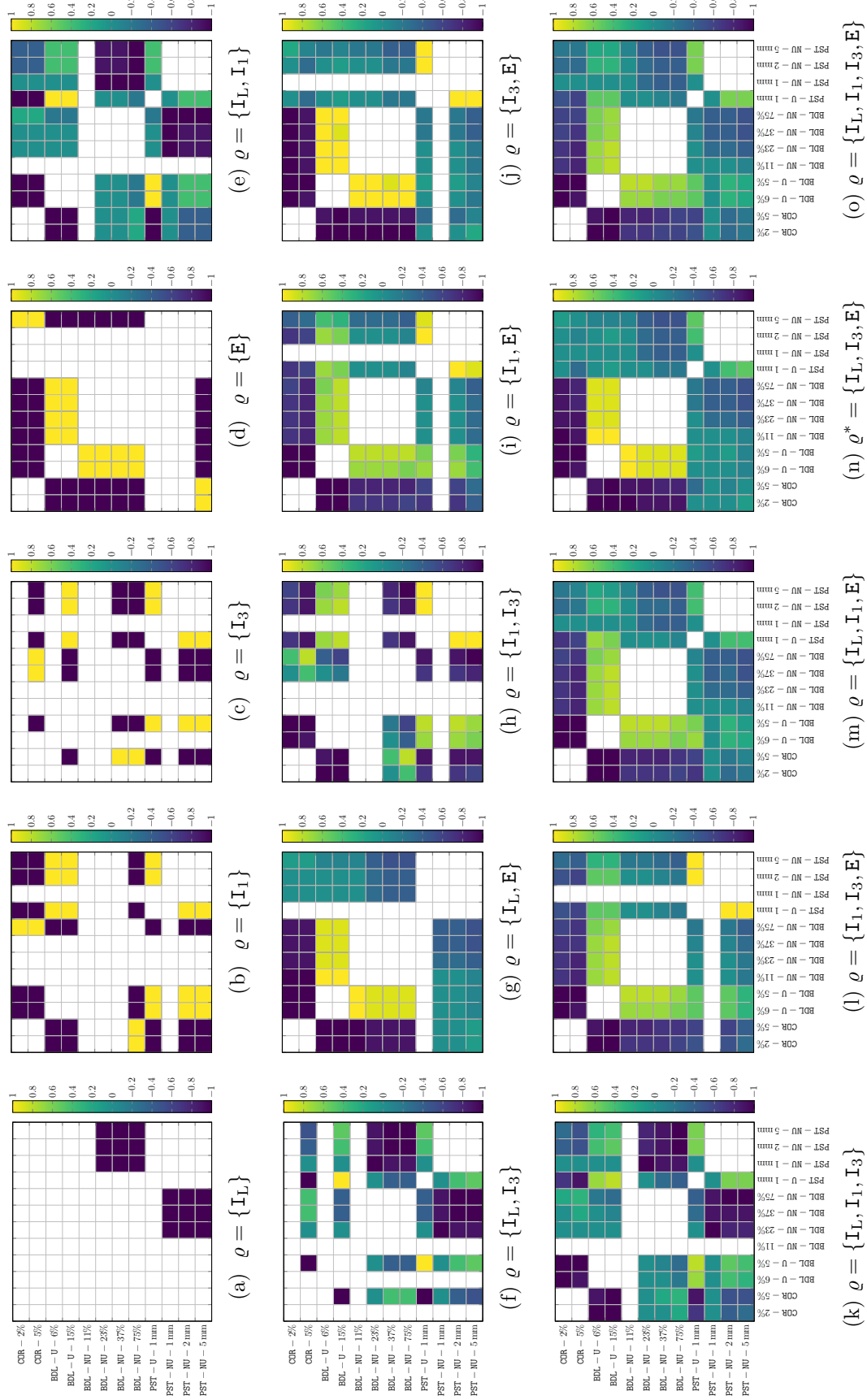
Figure 6.7 Illustration of the heat-maps for all measurement system configurations. The white cells represent either null or the cosine similarity values that are not considered. The simplified name of the sensors are: I_L: Lateral inclinometer, I_1: Longitudinal inclinometer 1, I_3: Longitudinal inclinometer 3, and E: Extensometer. Also, the abbreviations for the damage scenarios are COR: corrosion, BDL: bearing device lockage, PST: partial settlement, U: uniform, and NU: non-uniform.

i.e., $\varrho^* = \{\mathtt{I_L}, \mathtt{I_3}, \mathtt{E}\}$. Nevertheless, note that since the cosine similarity metric $\bar{p}$ is similar for the three sensor configurations, we expect in practice that they will have a similar damage distinguishability. The difference between the configuration shown in Figure 6.7m and the optimal one is the usage of the longitudinal inclinometer 1 instead of 3, i.e., using the sensor $\mathtt{I_1}$ instead of $\mathtt{I_3}$. In addition, the difference between the configuration shown in Figures 6.7o and the optimal one is the usage of the four sensors instead of three.

Figure 6.8 shows the cosine similarity matrices corresponding to the three configurations shown in Figures 6.7m-6.7o. The shaded lower triangle part of the matrices are the elements taken into account during this new calculation. Changing a measurement system configuration can alter the distinguishability such that it might increase for a subset of damage scenario pairs at the expense of its reduction for an another subset. In order to show this point numerically, we have re-calculated the average cosine similarity between the five damage types separately, rather than calculating it for all damage scenarios at once. For a pair of damage type, we consider the corresponding rows and columns, and take the average of their values. For instance, rows 5 to 8 and columns 3 and 4 form a new matrix that is associated with the non-uniform and uniform bearing device damage types pair. For this pair, the average cosine similarity metric $\bar{p}$ is equal to $+0.93$, $+0.73$, and $+0.70$. They respectively correspond to the matrices shown in Figures 6.8a to 6.8c.

For each measurement system configuration, we obtain ten values, each representing the distinguishability between a damage type pair as shown in Table 6.3. The first two columns identify the damage type pairs. The remaining columns are associated with the same configurations shown in Figures 6.7m-6.7o. The values in green correspond to the cases for which there is an improvement in terms of distinguishability for the configuration $\varrho = \{\mathtt{I_L}, \mathtt{I_3}, \mathtt{E}\}$ compared with the other two configurations. For the other damage type pairs, the distinguishability is either reduced or not changed.

From a practical point of view, once a cosine similarity value between a damage type pair $[\mathbf{P}]_{ij} \leq 0$, the pair is easily distinguishable. This is because the corresponding damage vectors are either in an opposite direction for $[\mathbf{P}]_{ij} < 0$, or they are perpendicular for $[\mathbf{P}]_{ij} = 0$. Therefore, we expect that most of the average cosine similarity metrics presented in Table 6.3 will result in a similar distinguishability. However, one particular case in this table is the uniform and non-uniform bearing device lockage damage type pair as identified by the red color for the three configurations. Using the longitudinal inclinometer 3 instead of 1, i.e., $\mathtt{I_3}$ instead of $\mathtt{I_1}$, makes the damage type pair indistinguishable, while other pairs show similar distinguishability. This situation is also shown by the yellow areas in Figures 6.7m-6.7o. Moreover, the comparison between using the configuration $\{\mathtt{I_L}, \mathtt{I_1}, \mathtt{E}\}$ and $\{\mathtt{I_L}, \mathtt{I_1}, \mathtt{I_3}, \mathtt{E}\}$ shows

(a) $\varrho^* = \{\mathtt{I_L}, \mathtt{I_3}, \mathtt{E}\}$

	1	2	3	4	5	6	7	8	9	10	11	12
1	1											
2	0.9	1										
3	−1	−0.9	1									
4	−1	−1	1	1								
5	−1	−0.9	1	1	1							
6	−0.9	−0.9	0.9	0.9	0.9	1						
7	−0.9	−0.8	0.9	0.9	0.9	1	1					
8	−0.9	−0.8	0.9	0.9	0.9	1	1	1				
9	0	−0.3	0	0.1	0	0	−0.2	−0.2	1			
10	0	0	0	0	0	−0.3	−0.4	−0.4	0	1		
11	0	−0.1	0	0.1	0	−0.3	−0.4	−0.4	0.4	0.9	1	
12	0.1	0	−0.1	−0.1	−0.1	−0.4	−0.5	−0.5	0.5	0.9	1	1

(b) $\varrho = \{\mathtt{I_L}, \mathtt{I_1}, \mathtt{E}\}$

	1	2	3	4	5	6	7	8	9	10	11	12
1	1											
2	1	1										
3	−1	−1	1									
4	−1	−1	1	1								
5	−0.7	−0.8	0.7	0.8	1							
6	−0.7	−0.8	0.7	0.8	0.9	1						
7	−0.7	−0.8	0.7	0.8	0.9	1	1					
8	−0.6	−0.7	0.6	0.7	0.9	1	1	1				
9	−0.7	−0.6	0.7	0.6	0	0	0	−0.1	1			
10	0	0	0	0	0	−0.3	−0.4	−0.4	0	1		
11	−0.3	−0.2	0.3	0.3	0	−0.3	−0.3	−0.4	0.4	0.9	1	
12	−0.2	−0.1	0.2	0.1	−0.1	−0.4	−0.5	−0.5	0.4	0.9	1	1

(c) $\varrho = \{\mathtt{I_L}, \mathtt{I_1}, \mathtt{I_3}, \mathtt{E}\}$

	1	2	3	4	5	6	7	8	9	10	11	12
1	1											
2	0.9	1										
3	−1	−0.9	1									
4	−1	−1	1	1								
5	−0.7	−0.8	0.7	0.8	1							
6	−0.7	−0.8	0.7	0.8	0.9	1						
7	−0.6	−0.7	0.6	0.7	0.9	1	1					
8	−0.6	−0.6	0.6	0.7	0.9	1	1	1				
9	−0.5	−0.6	0.5	0.5	0	0	−0.1	−0.2	1			
10	0	0	0	0	0	−0.3	−0.4	−0.4	0	1		
11	−0.3	−0.3	0.3	0.3	0	−0.3	−0.4	−0.4	0.6	0.8	1	
12	−0.2	−0.2	0.2	0.2	−0.1	−0.4	−0.5	−0.5	0.6	0.8	1	1

(d) Damage scenarios

ID	Type	Abbreviation
1	Corrosion	$\mathtt{COR} - 2\%$
2		$\mathtt{COR} - 5\%$
3	Uniform bearing	$\mathtt{BDL} - \mathtt{U} - 5.4\%$
4	device lockage	$\mathtt{BDL} - \mathtt{U} - 14.6\%$
5	Non-uniform	$\mathtt{BDL} - \mathtt{NU} - 10.2\%$
6	bearing device	$\mathtt{BDL} - \mathtt{NU} - 22.2\%$
7	lockage	$\mathtt{BDL} - \mathtt{NU} - 36.3\%$
8		$\mathtt{BDL} - \mathtt{NU} - 74.1\%$
9	Uniform partial settlement	$\mathtt{PST} - \mathtt{U} - 1\,\mathrm{mm}$
10	Non-uniform	$\mathtt{PST} - \mathtt{NU} - 1\,\mathrm{mm}$
11	partial settlement	$\mathtt{PST} - \mathtt{NU} - 2\,\mathrm{mm}$
12		$\mathtt{PST} - \mathtt{NU} - 5\,\mathrm{mm}$

Figure 6.8 Cosine similarity matrices. The lower part of the matrix that are shaded are considered in evaluating $\bar{p}$. For a better illustration, we use the damage scenario IDs instead of their abbreviation as presented in Table 6.2. The simplified name of the sensors are: $\mathtt{I_L}$: Lateral inclinometer, $\mathtt{I_1}$: Longitudinal inclinometer 1, $\mathtt{I_3}$: Longitudinal inclinometer 3, and $\mathtt{E}$: Extensometer.

that the distinguishability remains the same from both numerical and practical perspectives, while the latter configuration uses a larger number of sensors.

For the purpose of instrumenting all the spans from the Jacques Cartier Bridge, we suggest to use the configuration consisting in the lateral inclinometer, longitudinal inclinometer 1,

Table 6.3 The average cosine similarity $\bar{p}_{i,j}$ between the five damage types corresponding to the three measurement system configurations. The green values indicate an improvement in distinguishability between damage types. The simplified name of the sensors are: I_L: Lateral inclinometer, I_1: Longitudinal inclinometer 1, I_3: Longitudinal inclinometer 3, and E: Extensometer. Also, COR: corrosion, BDL: bearing device lockage, PST: partial settlement, U: uniform, and NU: non-uniform.

Damage type pair (i,j)		Measurement system configuration ϱ		
i	j	$\{I_L, I_3, E\}$	$\{I_L, I_1, E\}$	$\{I_L, I_1, I_3, E\}$
COR	BDL − U	−0.98	−1.00	−0.98
COR	BDL − NU	−0.89	−0.73	−0.69
COR	PST − U	−0.15	−0.65	−0.55
COR	PST − NU	+0.00	−0.13	−0.17
BDL − U	BDL − NU	+0.93	+0.73	+0.70
BDL − U	PST − U	+0.05	+0.65	+0.50
BDL − U	PST − NU	−0.02	+0.15	+0.17
BDL − NU	PST − U	−0.10	−0.03	−0.08
BDL − NU	PST − NU	−0.31	−0.30	−0.31
PST − U	PST − NU	+0.30	+0.27	+0.40

and the extensometer, i.e., $\varrho = \{I_L, I_1, E\}$ on each span. Such a configuration is expected to be able to detect the damages induced from a corrosion of $\geq 2\%$, a uniform bearing device lockage of $\geq 5.4\%$, a non-uniform bearing device lockage of $\geq 10.2\%$, and uniform and non-uniform partial settlements of $\geq 1\,\mathrm{mm}$. The damage detectability of the proposed configuration is based on the corresponding agents presented in Table 4.7.

Table 6.4 summarizes the two macroscopic criteria for the recommended sensor configuration; the third column shows the first criterion that is the minimum detectable anomaly magnitude $|h_a|_{\min}$. The last column shows the second criterion which is the minimum detectable anomaly growth rate. For instance, for the agent corresponding to the lateral inclinometer with the

Table 6.4 Minimum detectable anomaly magnitudes $|h_a|_{\min}$ and the minimum detectable anomaly growth rates ψ in years as a function of $|h_a|_{\min}$, i.e., $\psi = |h_a|_{\min}$ per N_ψ years. The shorthanded name of the sensors are I_L for the lateral inclinometer, I_1 for the longitudinal inclinometer on the axis 1, and E for the extensometer.

Sensor ID	ζ_{FP}	$|h_a|_{\min}$	N_ψ [year]
I_L	0.05	0.0017 °d	7
	0.10	0.0020 °d	8
I_1	0.05	0.0013 °d	6
	0.10	0.0013 °d	6
E	0.05	0.0100 mm	7
	0.10	0.0090 mm	6

annual false alarm ratio of $\zeta_{\text{FP}} = 0.05/\text{year}$, the minimum anomaly magnitude must be equal or greater than $0.0017°\text{d}$ per $\mathbb{N}_\psi = 7$ years.

6.4   Conclusion

This chapter introduced methodologies in order to associate an anomaly's magnitude to the severity levels of structural damages and optimize sensor configurations. Three structural damage types, namely corrosion, bearing device lockage, and partial settlement were modeled in order to study the effect of damage scenarios on structural responses. Furthermore, the minimum detectable anomaly magnitudes obtained in Chapter 4 are used along with the damage scenarios. This chapter used the cosine similarity method in order to introduce a metric for representing the distinguishability between damage scenarios using multiple sensors. From this metric, we identified that using the configuration consisting in the lateral inclinometer, longitudinal inclinometer 3, and the extensometer resulted in the maximum distinguishability across all damage scenarios. In addition, we observed that using a larger number of sensors does not guarantee a higher distinguishability. We found that using the configuration consisting in the lateral inclinometer, longitudinal inclinometer 1, and the extensometer resulted in the better distinguishability across different damage types. Therefore, we suggest to use this sensor configuration on each span of the Jacques Cartier Bridge. This sensor configuration along with the agents presented in Chapter 4 are able to detect anomalies with the minimum magnitude $|h_a|_{\min}$ and the minimum growth rate ψ according to Table 6.4. In addition, such a configuration is expected to detect the structural damages caused by a corrosion of $\geq 2\%$, a uniform bearing device lockage of $\geq 5.4\%$, a non-uniform bearing device lockage of $\geq 10.2\%$, and uniform and non-uniform partial settlements of $\geq 1\,\text{mm}$.

CHAPTER 7   Conclusion

This thesis proposed methodologies for detecting anomalies as well as for quantifying the *anomaly detectability* and the structural *damage distinguishability*, and applied those to the case study of the Jacques Cartier Bridge. The following sections provide the key aspects from the methodologies developed, the conclusions emanating from the analyses, the limitations encountered using the proposed methodologies, and the future possible research directions in order to address these limitations.

7.1   Thesis Conclusions

In the context of structural health monitoring where anomalies are rare events and observed responses are influenced by external effects, rightfully detecting anomalies plays a key role within the process of quantifying the anomaly detectability; although detecting anomalies has been studied in the literature, these studies have not accounted for an anomaly's characteristics including its magnitude and duration. As a result, existing methods are not able to quantify probabilistically the anomaly detectability with respect to anomalies' characteristics as well as other factors such as the time required to rightfully detect an anomaly and the number of false alarms that a decision maker expects to encounter. In addition, existing methods are not able to associate the magnitudes of detected anomalies to the severity levels of a wide range of structural damages. In order to address these limitations as well as other limitations identified in the literature, this research work presented an anomaly detection procedure based on Bayesian dynamic linear models (BDLMs) along with the reinforcement learning (RL) and imitation learning (IL) approaches.

The formulation of BDLM relies on assembling generic components in order to estimate the degradation's kinematics consisting in structural degradation, degradation's speed, and degradation's acceleration. In addition, these components account for the external effects such as the temperature to remove them from the measurements. As a result, a BDLM allows decomposing the measurements collected from the sensors in order to obtain the underlying stationary and non-stationary structural degradation's speeds as well as their probabilities. We employed BDLMs on the measurements collected from the inclinometers, extensometers, and thermometers installed on the Jacques Cartier Bridge. The analysis of the sensors determined the stationary and non-stationary structural responses and their probabilities, which are the two criteria during the process of detecting anomalies as it will be explained later in this section. We observed that relying solely on the probabilities of the non-

stationary structural degradation's speed is not sufficient for detecting anomalies; There are situations for which such probabilities are not high, while the corresponding speeds indicate the presence of anomalies, and therefore, a decision maker requires to take into account such speeds.

This research work proposed a stochastic anomaly function that models an anomaly based on its magnitude, its duration, and the time of occurrence, which have a key role within the process of detecting anomalies and quantifying the anomaly detectability of the sensors; Used along with the empirical models obtained from BDLM, we were able to generate abnormal synthetic time series and the subsequent environment's state-space consisting in stationary and non-stationary structural degradation's speeds as well as their probabilities. The resulting environment, with which an agent interact, mimics the degradation's kinematics associated with the real measurements, and accounts for the structural degradation's speeds and their probabilities as the two sources of information for decision making.

Following the simulation of such an environment, we formulated the problem of detecting anomalies as a sequential decision making problem for which an agent is required to take the action within an environment in order to trigger an alarm when an anomaly is present, and not to trigger an alarm, otherwise. We proposed two methods for treating this problem: The first method relies on the reinforcement learning approach, while the second one uses the imitation learning approach as a basis for training an agent. The former uses a reward function as a feedback signal in order to encourage or prevent an agent to take an action with regards to the time to trigger an alarm, while the latter uses a demonstration set in order to imitate an expert's alarm triggering policy. In addition, in order to quantify the anomaly detectability for each sensor, we proposed two metrics, namely the *annual false positive ratio*, and the *probability of true positive detections*. The former identifies the number of false alarms per year that a decision maker expects to encounter. The latter quantifies the probability of rightfully detecting anomalies as a function of three variables; The first two variables are the anomaly's magnitude and its duration that are accessible from the simulated environment. The third variable is the time required by an agent to detect an anomaly after its starting time.

We employed both anomaly detection methods for each sensor type on the Jacques Cartier Bridge in order to obtain alarm triggering policies for the target annual false positives of one per 10 and 20 years. The resulting alarm triggering policies were examined on real time series measurements obtained from the Jacques Cartier Bridge. We showed that the anomaly detection for both annual false positive ratios successfully identified anomalies earlier than the time when a stationary regime switches to a non-stationary one. A key aspect found is

that the reward values control the annual false positive ratios: setting a higher true positive reward value for rightfully detecting anomalies encourages an agent to trigger an alarm more often, which results in more false alarms. On the other hand, assigning a higher negative reward value for wrongfully detecting anomalies prevents an agent for triggering an alarm, and subsequently we to obtain a lower annual false positive ratio.

While succeeding at detecting anomalies using the reinforcement learning method, we showed that the imitation learning approach requires less data to train an agent, resulting in lower computational costs. We examined different alarm triggering policies using imitation learning for the target annual false positive ratios of one per 5, 10, 20, and 40 years. It was found that, in most cases, the resulting decision boundaries are very close to each other showing a potential lack of robustness in achieving a target annual false positive ratio.

Furthermore, we used the resulting alarm triggering policies in order to evaluate the probability of true positive detections. In addition, we reported the 90% probability of true positive detections in the form of isosurfaces for the inclinometers and extensometers for both annual false positive ratios of one false alarm per 10 and 20 years, and for both anomaly detection methods. We extracted two macroscopic metrics from these isosurfaces, namely the *minimum detectable anomaly magnitude* and the *minimum detectable anomaly growth*; These two metrics are another key aspect to enable quantifying the capacity of each sensor in detecting anomalies. For both anomaly detection methods, we showed that there is a tradeoff between the annual false positive ratio and the probability of true positive detections: Lowering the annual false positive ratio comes at the expense of a decrease in the probability of true positive detections. Moreover, the results showed that both reinforcement and imitation learning based anomaly detection methods yield similar anomaly detectability capacities.

Following the quantification of the anomaly detectability, we proposed methodologies based on physics-based models and the cosine similarity metric for quantifying the structural damage distinguishability, which aims at associating an anomaly's magnitude to the severity levels of different structural damages. To this end, the current research work used a physics-based model in order to simulate structural damages and associate their severity levels to the minimum detectable anomaly of sensors. We presented three types of structural damages with various severity; they are corrosion, bearing device lockages, and partial settlements. Furthermore, the current work used the cosine similarity metric in order to measure the capacity of different sensor configurations at distinguishing damage scenarios and chose the configuration that maximizes the distinguishability between them. We showed that using the configuration consisting in the lateral inclinometer, longitudinal inclinometer on the axis 1, and the extensometer resulted in the best distinguishability across different damage types.

This sensor configuration along with the agents presented in Chapter 4 are able to detect anomalies with the minimum magnitude $|h_a|_{\min}$ and the minimum growth rate ψ reported in Table 6.4. In addition, such a configuration is expected to detect the structural damages caused by a corrosion of $\geq 2\%$, a uniform bearing device lockage of $\geq 5.4\%$, a non-uniform bearing device lockage of $\geq 10.2\%$, and uniform and non-uniform partial settlements of $\geq 1\,\text{mm}$.

In conclusion, the methodologies proposed in this thesis showed the capacity of reinforcement and imitation learning approaches to successfully quantify: (i) the anomaly detectability of different sensor types with respect to the target annual false positive ratio as well as the probability of true positive detections, and (ii) the structural damage distinguishability in order to associate an anomaly's magnitude to the severity levels of structural damages, and to help determine the sensor configuration that maximizes the distinguishability between these damages. These methods showed a good performance on different measurements collected on the Jacques Cartier Bridge; They determined the minimum anomaly detectability, and quantified the damage detectability of each sensor. Overall, the proposed methods introduced a novel approach towards detecting anomaly, quantifying the anomaly detectability, and the structural damage distinguishability, which laid the foundation for further applications for other civil infrastructures in the context of long-term structural health monitoring.

7.2 Limitations

In this section we present the limitations of the methodologies developed in the current research work regarding detecting anomalies, quantifying the anomaly detectability, and quantifying the damage detectability. Resolving these limitations would further improve the scalability and robustness of the proposed methodologies.

7.2.1 Environment's State-Space

In this thesis, we used two explanatory variables forming a two-dimensional environment's state-space for the decision making process. They are the structural degradation's speeds and the probability of the non-stationary regime. In addition, we discretized the environment when we applied reinforcement learning approach for detecting anomalies as described in Section 4.1.1 in Chapter 4. However, there are two limitations when using such an environment for with which an agent interacts; First, determining a grid size for the discretization purpose relies on a grid search in the sense that we need to try different grid size configurations in order to ensure the convergence of the Q-values [85]. Second, using other explanatory variables

in the decision making process has not been incorporated. Besides, increasing the number of explanatory variables in simulating an environment increases the number of discretized regions with $\mathcal{O}(\mathbb{N}^{\mathbb{S}})$, where $\mathbb{S}$ is the number of variables, and $\mathbb{N}$ is the number of discretization in each dimension assuming we use an equal number of divisions. As a result, the computational costs for training an agent becomes more demanding. One potential solution to relax these limitations, when using the reinforcement learning approach, is to use function approximation methods presented in Section 2.3.1 in Chapter 2; These methods approximate the action-value function using an approximator [61, 62], so there is no need to discretize the environment's state space. In addition, advanced function approximators based on using deep neural networks [96] can cope with an increase in the number of dimensions, and hence enable the anomaly detection procedure to account for more explanatory variables.

The next limitation of the current research work is related to the capacity to interpret data without interruption after anomalies are detected. The challenge lies in the fact that structural anomalies introduce non-stationary regimes in structural responses. Such a change in the patterns requires adapting the reinforcement learning structure for operating in a non-stationary environment. The full-scale operation of a sensor network and of the data analysis associated with it require being able to operate without interruption even in the case where anomalies are identified, either following a change in the condition of the structure or following planned events such as maintenance activities. A possible solution to relax such a limitation is to design additional reward features and allow an agent to keep interacting with the environment while collecting these reward values.

Another limitation of the current research is related to using a grid $d_{\mathbb{M}\times\mathbb{N}}$ for downsampling presented in Chapter 5 in regard with the imitation learning approach. Here, we used the same grid that were using in discretizing an environment in the reinforcement learning approach. This grid is a non-uniform grid with respect to the probability of the non-stationary regime $\pi^{\mathbb{ns}}$; The grid size for the probabilities $\pi^{\mathbb{ns}} < 0.05$ differs from the probabilities $\pi^{\mathbb{ns}} > 0.05$. When drawing samples for the issues region in the imitation learning approach, such a difference results in changing the density of the acquired samples and hence their probability samples. In order to overcome this limitation and improve the results, we suggest to use a uniform grid for the entire environment's domain and draw samples from, or employ algorithms such a DAgger [80], explained in Section 2.3.2, which copes with the non-uniform data sampling issue.

Another limitation of the current research relies on the nature of the environment. The anomaly detection problem was formulated using a Markov decision process (MDP) hypothesizing that an agent observes a true state at each time step; The structural degradation's

speed is a hidden-state variable which is obtained from an empirical model and represented by a normal distribution with a mean μ_t^{T} and a standard deviation σ_t^{T} for each time step. However, we simulated the environment using its expected values assuming they represent the true degradation's speed. One solution to relax such a limitation is to formulate the anomaly detection problem as a *partially observable Markov decision process* (POMDP). In this configuration, an agent observes the states with uncertainty. As a result, the policy $\pi(p(\mathbf{s}))$ is a function of the probability of the state $\mathbf{s}$, rather than being a function of the state in the case of MDP, i.e., $\pi(\mathbf{s})$. Solving problems in a POMDP setup is more demanding and require to use approximation methods [97]. The reader is referred to specialized literature such as Kaelbling et al. [98] and Hauskrecht [97] for detailed review of the subject matter.

7.2.2 Structural Damage Scenarios and Detectability

The damage scenarios presented in the current research work are limited to three types. However, a structure could experience other types of damages or there might be specific damages that are of greatest interest for a decision maker. In addition, the proposed structural damages in this work represent global damages and the local damages were not considered; For instance, in the case of corrosion, we considered the effect of the corrosion on all the steel elements of a span neglecting the situation for which the corrosion would occur on a part of a span. Moreover, the damage detectability presented in Chapter 6 is deterministic in the sense that it only identifies the possibility of certain structural damages. However, such a diagnosis capacity is not quantified probabilistically such that for a detected anomaly we would be able to assign a probability value for the possible causes. A solution to address these limitations is to build a joint observations set from the various structural damages, both locally and globally, and employ the Bayes' theorem in order to infer the probability of these damages for when an agent detects anomalies.

7.3 Future Work

This thesis laid the foundation for detecting anomalies in the context of sequential decision making, and presented the anomaly detectability as well as the structural damage distinguishability. This section presents future research directions, which include potential improvements on the anomaly detection method as well as the structural damage detectability.

7.3.1 Detecting Anomalies Using Additional Explanatory Variables

Chapters 4 and 5 showed that using the two explanatory variables improves the decision making with regard to detecting anomalies compared with solely relying on the probability of the non-stationary regimes obtained from the Bayesian dynamic linear models. We suggest to use additional explanatory variables to examine the anomaly detectability of an agent. These additional variables potentially rely on the *relative structural degradation vector* $\hat{\mathbf{x}}^{\mathrm{L}}$ defined as

$$\hat{\mathbf{x}}_t^{\mathrm{L}} = [(x_t^{\mathrm{L}} - x_{t-1}^{\mathrm{L}}) \ (x_t^{\mathrm{L}} - x_{t-2}^{\mathrm{L}}) \ \ldots \ (x_t^{\mathrm{L}} - x_{t-k}^{\mathrm{L}})]^\intercal, \ t \geq k, \tag{7.1}$$

where, at a time t, the vector $\hat{\mathbf{x}}^{\mathrm{L}}$ consists in k consecutive relative structural degradation values corresponding to the previous time steps. For instance for $k = 1$, the relative structural degradation vector reduces to an explanatory variable representing the difference between the current degradation and the one associated with the previous time step. i.e, $\hat{\mathbf{x}}^{\mathrm{L}} = x_t^{\mathrm{L}} - x_{t-1}^{\mathrm{L}}$. In this setup, an agent takes into account the changes in the degradation patterns for the decision making process with regard to triggering an alarm. Note that the resulting environment's state-space $\mathcal{S} \in \mathbb{R}^{k+2}$ is a $(k+2)$-dimensional space, where the number 2 corresponds to the two explanatory variables presented in this thesis.

7.3.2 Enhancing Detecting Anomalies Using Alternative Learning Algorithms

In this work, detecting anomalies relies on using either the reinforcement learning approach in a discretized environment, or the imitation learning approach performed on a demonstration set obtained from the expert's policy. However, the anomaly detection could be enhanced in two ways; The first improvement is related to employing the reinforcement learning approaches based on the function approximation, in particular, the ones that are built using deep neural networks. These methods learns the optimal policy in a continues environment's state-space, so there is no need for discretizing an environment. In addition, they are able to handle high dimensional environments enabling an agent to account for high number of explanatory variables such as the aforementioned relative structural degradation vector.

The second improvement is associated with using both reinforcement and imitation learning approaches in a same framework. The imitation learning approach presented in Chapter 5 is, on one hand, faster by orders of magnitudes compared with the reinforcement learning approach presented in Chapter 4. On the other hand, the imitation learning approach is not able to provide information about how the expert makes decisions with regard to triggering anomalies, and it is prone to suboptimal policies provided by an expert [82, 99]. We suggest to apply both reinforcement and imitation learning approaches in a same framework

in order to improve the learning process. When using the reinforcement learning approach in a discretized environment, the initial Q-values could be estimated by allowing an agent to imitate the expert's policy. To this end, the imitation learning method can be used to learn the optimal policy for each discretized region. The corresponding Q-values are then estimated using an independent test dataset containing both normal and abnormal episodes. Afterwards, a reinforcement learning agent learns the optimal policy using these initialized Q-values. The same procedure also can be used using the reinforcement learning approach in a continuous environment; when using a deep neural network, the initial weights and biases can be estimated using the optimal policy obtained from the imitation learning approach. Using both reinforcement learning and imitation learning approaches has been examined in the literature. For instance, Minh et al. [61] used such an approach in the game of Go: they used demonstrations obtained from the expert Go players to train an agent in the context of the imitation learning approach, and then applied the reinforcement learning method in order to improve the optimal policy.

7.3.3   Detecting Anomalies in an Uncertain Environment's State-Space

The anomaly detection problem presented in Chapter 4 was formulated in a Markov decision process (MDP) setup. As discussed in the limitation section, this is not an accurate setup, because we only used the expected values of the degradation's speed without taking into account uncertainty around these values. In order to account for the uncertainty, we suggest to formulate the anomaly detection problem as a partially observable Markov decision process (POMDP). As such, the resulting environment is presented with uncertainty for which an agent observe the states with an uncertainty. Although solving POMDP problems in the context of decision making is more demanding compared with a MDP [97], employing reinforcement learning methods based on a Bayesian neural network [96] can address such an issue. These methods have been developed to take uncertain states as an input and approximate the Q-values in a probabilistic manner.

7.3.4   Probabilistic Models for the Damage Detectability

The diagnostic capacity of the damage detectability presented in Chapter 6 results in only deterministically associating the magnitude of a detected anomaly to a set of structural damages with different severity levels. This implies that it is not able to quantify the cause of each damage within this set, probabilistically. Besides, the occurrence of different structural damages depends on the characteristics of a civil infrastructure. For instance, it is possible that the corrosion causes an anomaly with a higher probability for a bridge with a steel truss

deck compared with a concrete one. Therefore, the uncertainty of the detected damages ought to be quantified in order to improve the damage detectability. To this end, we suggest to build a joint observations set from various structural damages, and employ the Bayes' theorem in order to infer the probability of these damages for when an agent detects them.

REFERENCES

[1] J.-A. Goulet, *Probabilistic machine learning for civil engineers.* The MIT Press, 2020.

[2] L. H. Nguyen and J.-A. Goulet, "Anomaly detection with the switching Kalman filter for structural health monitoring," *Structural Control and Health Monitoring*, vol. 25, no. 4, p. e2136, 2018.

[3] "Atari breakout," Yandex LLC, Tech. Rep., 2022.

[4] "Jacques Cartier Bridge technical data sheet," Pont Jacques Cartier and Champlain Inc., Tech. Rep., 2022.

[5] "Jacques Cartier Bridge deck replacement: Drawing no. 125212-220001," Groupement SNC-Lavalin, Tech. Rep., 2002.

[6] "Jacques Cartier Bridge miscellaneous steel repairs: Drawing no. 125138-07," Groupement SNC-Lavalin, Tech. Rep., 1999.

[7] M. Sánchez-Silva and J. Riascos-Ochoa, "15 - seismic risk models for aging and deteriorating buildings and civil infrastructure," 2013, pp. 387–409.

[8] S. Démurger, "Infrastructure development and economic growth: An explanation for regional disparities in china?" *Journal of Comparative Economics*, vol. 29, no. 1, pp. 95–117, 2001.

[9] "Collapse of I-35W highway bridge," National Transportation Safety Board, Tech. Rep., 2008.

[10] L. Balsamo, R. Betti, and H. Beigi, "A structural health monitoring strategy using cepstral features," *Journal of Sound and Vibration*, vol. 333, no. 19, pp. 4526 – 4542, 2014.

[11] W. Ostachowicz, R. Soman, and P. Malinowski, "Optimization of sensor placement for structural health monitoring: A review," *Structural Health Monitoring*, vol. 18, no. 3, pp. 963–988, 2019.

[12] D. Agdas, J. A. Rice, J. R. Martinez, and I. R. Lasa, "Comparison of visual inspection and structural health monitoring as bridge condition assessment methods," *Journal of Performance of Constructed Facilities*, vol. 30, no. 3, p. 04015049, 2016.

[13] C. R. Farrar and K. Worden, *An introduction to structural health monitoring.* Springer, 2010.

[14] "Monitoring the state of Canada's core public infrastructure: The Canadian infrastructure report card 2019," BluePlan Engineering, Tech. Rep., 2019.

[15] M. George and C. Sekine, "Undrestanding Canada's infrastructure crisis and constrction trends," Trisura, Tech. Rep., 2017.

[16] D. Stiff and P. Smetanin, "Public infrastructure underinvestment: The risk to Canada's economic growth," Risk Analytica, Tech. Rep., 2010.

[17] A. P. Black, "2022 bridge report," American Road and Transportation Builder Association, Tech. Rep., 2022.

[18] "2021 report card for America's infrastructure," American Society of Civil Engineers, Tech. Rep., 2021.

[19] C. Farrar and K. Worden, *Structural health monitoring: A machine learning perspective.* John Wiley and Sons, 2013.

[20] O. C. Zienkiewicz, R. L. Taylor, and J. Z. Zhu, *The Finite Element Method: Its Basis and Fundamentals.* Butterworth-Heinemann, 2013.

[21] M. I. Friswell, *Damage Identification using Inverse Methods.* Springer, 2008.

[22] B. Jaishi and W.-X. Ren, "Damage detection by finite element model updating using modal flexibility residual," *Journal of Sound and Vibration*, vol. 290, no. 1, pp. 369–387, 2006.

[23] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine learning for anomaly detection: A systematic review," *IEEE Access*, vol. 9, pp. 78 658–78 700, 2021.

[24] M. Yu and S. Sun, "Policy-based reinforcement learning for time series anomaly detection," *Engineering Applications of Artificial Intelligence*, vol. 95, p. 103919, 2020.

[25] M. Morgantini, R. Betti, and L. Balsamo, "Structural damage assessment through features in quefrency domain," *Mechanical Systems and Signal Processing*, vol. 147, p. 107017, 2021.

[26] A. Achilli, G. Bernagozzi, R. Betti, P. Diotallevi, L. Landi, S. Quqa, and E. Tronci, "On the use of multivariate autoregressive models for vibration-based damage detection and localization," *Smart Structures and Systems*, vol. 27, pp. 335–350, 2021.

[27] Y. Ying, J. Garrett, I. Oppenheim, L. Soibelman, J. Harley, J. Shi, and Y. Jin, "Toward data-driven structural health monitoring : Application of machine learning and signal processing to damage detection," *Journal of Computing in Civil Engineering*, vol. 27, pp. 667–680, 2013.

[28] R. Mohammadi Ghazi and O. Büyüköztürk, "Damage detection with small data set using energy-based nonlinear features," *Structural Control and Health Monitoring*, vol. 23, no. 2, pp. 333–348, 2016.

[29] M. Malekzadeh, G. Atia, and F. N. Catbas, "Performance-based structural health monitoring through an innovative hybrid data interpretation framework," *Journal of Civil Structural Health Monitoring*, vol. 5, no. 3, pp. 287–305, 2015.

[30] L. Bull, K. Worden, G. Manson, and N. Dervilis, "Active learning for semi-supervised structural health monitoring," *Journal of Sound and Vibration*, vol. 437, pp. 373–388, 2018.

[31] L. Li, M. Morgantini, and R. Betti, "Structural damage assessment through a new generalized autoencoder with features in the quefrency domain," *Mechanical Systems and Signal Processing*, vol. 184, p. 109713, 2023.

[32] J.-A. Goulet, "Bayesian dynamic linear models for structural health monitoring," *Structural Control and Health Monitoring*, vol. 24, no. 12, p. e2035, 2017.

[33] O. Avci, O. Abdeljaber, S. Kiranyaz, M. Hussein, M. Gabbouj, and D. J. Inman, "A review of vibration-based damage detection in civil structures: From traditional methods to machine learning and deep learning applications," *Mechanical Systems and Signal Processing*, vol. 147, p. 107077, 2021.

[34] W.-H. Hu, D.-H. Tang, J. Teng, S. Said, and R. Rohrmann, "Structural health monitoring of a prestressed concrete bridge based on statistical pattern recognition of continuous dynamic measurements over 14 years," *Sensors*, vol. 18, p. 4117, 2018.

[35] K. P. Murphy, *Machine learning : A probabilistic perspective.* The MIT Press, 2013.

[36] I. Farreras-Alcover, M. K. Chryssanthopoulos, and J. E. Andersen, "Regression models for structural health monitoring of welded bridge joints based on temperature, traffic and strain measurements," *Structural Control and Health Monitoring*, vol. 14, no. 6, pp. 648–662, 2015.

[37] E. P. Carden and J. M. Brownjohn, "ARMA modelled time-series classification for structural health monitoring of civil infrastructure," *Mechanical Systems Signal Processing*, vol. 22, no. 2, pp. 295 – 314, 2008.

[38] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Processing*, vol. 99, pp. 215 – 249, 2014.

[39] S. Ahmad, A. Lavin, S. Purdy, and Z. Agha, "Unsupervised real-time anomaly detection for streaming data," *Neural Computation*, vol. 262, pp. 134–147, 2017.

[40] M. Gaur, S. Makonin, I. V. Bajić, and A. Majumdar, "Performance evaluation of techniques for identifying abnormal energy consumption in buildings," *IEEE Access*, vol. 7, pp. 62 721–62 733, 2019.

[41] S. W. Doebling, C. R. Farrar, and M. B. Prime, "A summary review of vibration-based damage identification methods," *Shock and Vibration Digest*, vol. 30, pp. 91–105, 1998.

[42] A. K. Chopra, *Dynamics of Structures: Theory and Applications to Earthquake Engineering*.   Prentice Hall, 1995.

[43] D. C. Kammer, "Sensor placement for on-orbit modal identification and correlation of large space structures," *Journal of Guidance, Control, and Dynamics*, vol. 14, no. 2, pp. 251–259, 1991.

[44] M. Sanayei, C. Dicarlo, P. Rohela, E. Miller, and M. Kilmer, "Sensor placement using fisher information matrix for robust finite element model updating," *Journal of Life Cycle Reliability and Safety Engineering*, vol. 4, 2015.

[45] F. Ghayem, B. Rivet, C. Jutten, and R. C. Farias, "Optimal sensor placement for signal extraction," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019.

[46] S. Soni, S. Das, and A. Chattopadhyay, "Optimal sensor placement for damage detection in complex structures," *Smart Materials, Adaptive Structures and Intelligent Systems*, vol. 2, pp. 565–571, 2009.

[47] Z. H. Qureshi, T. S. Ng, and G. C. Goodwin, "Optimum experimental design for identification of distributed parameter systems," *International Journal of Control*, vol. 31, no. 1, pp. 21–29, 1980.

[48] E. Heredia-Zavoni, R. Montes-Iturrizaga, and L. Esteva, "Optimal instrumentation of structures on flexible base for system identification," *Earthquake Engineering and Structural Dynamics*, vol. 28, no. 12, pp. 1471–1482, 1999.

[49] M. West and J. Harrison, *Bayesian Forecasting and Dynamic Models.* Springer, 1999.

[50] J.-A. Goulet and K. Koo, "Empirical validation of Bayesian dynamic linear models in the context of structural health monitoring," *Bridge Engineering*, vol. 23, no. 2, p. 05017017, 2018.

[51] L. H. Nguyen and J.-A. Goulet, "Structural health monitoring with dependence on non-harmonic periodic hidden covariates," *Engineering Structures*, vol. 166, pp. 187–194, 2018.

[52] L. H. Nguyen, I. Gaudot, S. Khazaeli, and J.-A. Goulet, "A kernel-based method for modeling non-harmonic periodic phenomena in Bayesian dynamic linear models," *Frontiers in Built Environment*, vol. 5, p. 8, 2019.

[53] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[54] H. A. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE transactions on Automatic Control*, vol. 33, no. 8, pp. 780–783, 1988.

[55] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning.* The MIT Press, 2016.

[56] S. Khazaeli, L. H. Nguyen, and J.-A. Goulet, "Anomaly detection using state-space models and reinforcement learning," *Structural Control and Health Monitoring*, vol. 28, no. 6, p. e2720, 2021.

[57] D. Agarwal, "Detecting anomalies in cross-classified streams: A Bayesian approach," *Knowledge and Information Systems*, vol. 11, pp. 29–44, 2006.

[58] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach.* Prentice Hall Press, 2009.

[59] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Machine Learning*, vol. 3, pp. 9–44, 1988.

[60] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with deep reinforcement learning," DeepMind Technologies, Tech. Rep., 2013.

[61] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, 2015.

[62] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction.* The MIT Press, 2018.

[63] R. E. Bellman, "A Markovian decision process," *Journal of Mathematics and Mechanics*, pp. 679–684, 1957.

[64] J. Tsitsiklis and B. Van Roy, "An analysis of temporal-difference learning with function approximation," *IEEE Transactions on Automatic Control*, vol. 42, no. 5, pp. 674–690, 1997.

[65] M. G. Lagoudakis and R. Parr, "Least-squares policy iteration," *Journal of Machine Learning Research*, vol. 4, pp. 1107–1149, 2003.

[66] B. Dai, A. Shaw, L. Li, L. Xiao, N. He, Z. Liu, J. Chen, and L. Song, "SBEED: Convergent reinforcement learning with nonlinear function approximation," *The 35th International Conference on Machine Learning*, vol. 80, pp. 1125–1134, 2018.

[67] C. J. C. H. Watkins, "Learning from delayed rewards," Ph.D. dissertation, Cambridge University, 1989.

[68] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming.* Athena Scientific, 1996.

[69] S. W. Carden, "Convergence of reinforcement learning algorithm in continuous domains," Ph.D. dissertation, Clemson University, 2014.

[70] E. Eyal and M. Yishay, "Learning rates for Q-learning," *Journal of Machine Learning Research*, vol. 5, pp. 1–25, 2004.

[71] M. Yogeswaran and S. G. Ponnambalam, "Reinforcement learning: Exploration–exploitation dilemma in multi-agent foraging task," *Operational Research*, vol. 49, no. 3, pp. 223–236, 2012.

[72] N. R. Sturtevant, "An analysis of UCT in multi-player games," in *Proceedings of the 6th international conference on Computers and Games*, 2008.

[73] S. Korkmaz, "A review of active structural control: Challenges for engineering informatics," *Computers and Structures*, vol. 89, no. 23, pp. 2113–2132, 2011.

[74] T. T. Soong and G. P. Cimellaro, "Future directions in structural control," *Structural Control and Health Monitoring*, vol. 16, no. 1, pp. 7–16, 2009.

[75] B. Adam and I. Smith, "Reinforcement learning for structural control," *Journal of Computing in Civil Engineering*, vol. 22, pp. 133–139, 2008.

[76] A. Khalatbarisoltani, M. Soleymani, and M. Khodadadi, "Online control of an active seismic system via reinforcement learning," *Structural Control and Health Monitoring*, vol. 26, no. 3, p. e2298, 2019.

[77] C. Huang, Y. Wu, Y. Zuo, K. Pei, and G. Min, "Towards experienced anomaly detector through reinforcement learning," in *Proceedings of the 32nd Conference on Artificial Intelligence*, 2018.

[78] M. Oh and G. Iyengar, "Sequential anomaly detection using inverse reinforcement learning," in *Proceedings of the 25th International Conference on Knowledge Discovery and Data Mining*, 2019, pp. 1480–1490.

[79] D. Pomerleau, "ALVINN: An autonomous land vehicle in a neural network," in *Proceedings of Neural Information Processing Systems*, 1989.

[80] S. Ross, G. Gordon, and D. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, vol. 15, 2011.

[81] A. Hussein, M. M. Gaber, E. Elyan, and C. Jayne, "Imitation learning: A survey of learning methods," *Computing Surveys*, vol. 50, 2017.

[82] S. Arora and P. Doshi, "A survey of inverse reinforcement learning: Challenges, methods and progress," *Artificial Intelligence*, vol. 297, p. 103500, 2021.

[83] L. Nguyen and J.-A. Goulet, "Structural health monitoring with dependence on non-harmonic periodic hidden covariates," *Engineering Structures*, vol. 166, pp. 187–194, 2018.

[84] I. Gaudot, L. H. Nguyen, S. Khazaeli, and J.-A. Goulet, "OpenBDLM, an open-source software for structural health monitoring using Bayesian dynamic linear models," in *13th International Conference on Applications of Statistics and Probability in Civil Engineering*, 2019.

[85] A. Geramifard, T. J. Walsh, S. Tellex, G. Chowdhary, J. P. How, and N. Roy, "A tutorial on linear function approximators for dynamic programming and reinforcement learning," *Foundations and Trends in Machine Learning*, vol. 6, no. 4, pp. 375– 454, 2013.

[86] N. D. Daw, *Neuroeconomics.* Academic Press, 2014.

[87] J.-A. Goulet, L. Nguyen, and S. Amiri, "Tractable approximate Gaussian inference for Bayesian neural networks," *Journal of Machine Learning Research*, vol. 22, no. 251, pp. 1–23, 2021.

[88] Z. Zheng, J. Oh, and S. Singh, "On learning intrinsic rewards for policy gradient methods," in *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[89] S. Singh, R. L. Lewis, A. G. Barto, and J. Sorg, "Intrinsically motivated reinforcement learning: An evolutionary perspective," *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 2, pp. 70–82, 2010.

[90] D. Abel, W. Dabney, A. Harutyunyan, M. Ho, M. Littman, D. Precup, and S. Singh, "On the expressivity of markov reward," in *Proceedings of Neural Information Processing Systems*, 2021.

[91] D. Rengarajan, G. Vaidya, A. Sarvesh, D. Kalathil, and S. Shakkottai, "Reinforcement learning with sparse rewards using guidance from offline demonstration," in *International Conference on Learning Representations*, 2022.

[92] M. J. Mataric, "Reward functions for accelerated learning," in *In Proceedings of the Eleventh International Conference on Machine Learning*, 1994.

[93] U. Syed and R. E. Schapire, "A reduction from apprenticeship learning to classification," in *Advances in Neural Information Processing Systems*, vol. 23, 2010.

[94] Y. Li, J. Song, and S. Ermon, "InfoGAIL: Interpretable imitation learning from visual demonstrations," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[95] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques.* Morgan Kaufmann, 2011.

[96] L.-H. Nguyen and J.-A. Goulet, "Analytically tractable hidden-states inference in bayesian neural networks," *Journal of Machine Learning Research*, vol. 23, no. 50, pp. 1–33, 2022.

[97] M. Hauskrecht, "Value-function approximations for partially observable markov decision processes," *Journal of Artificial Intelligence Research*, vol. 13, no. 1, pp. 33–94, 2000.

[98] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial Intelligence*, vol. 101, no. 1, pp. 99–134, 1998.

[99] J. Kober and J. Peters, "Imitation and reinforcement learning," *IEEE Robotics and Automation Magazine*, vol. 17, no. 2, pp. 55–62, 2010.

APPENDIX A   BDLM MODELS

Model Matrices for the Inclinometers

The model matrices $\{\mathbf{A}_t, \mathbf{C}_t, \mathbf{Q}_t, \mathbf{R}_t\}$ for the stationary and non-stationary model matrices for the inclinometers $\mathtt{I}$ are defined following

Stationary Model Matrices $\mathcal{M}^{\mathtt{s,obs}}$

$$
\mathbf{A}_t^{\mathtt{s}} = \text{block diag} \left(\begin{bmatrix} 1 & \Delta t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & \tilde{\boldsymbol{k}}_t^{\mathtt{KR,I}} \\ \mathbf{0} & \mathbf{I}_{10\times 10} \end{bmatrix}, \phi^{\mathtt{AR,I}}, 1, \begin{bmatrix} 0 & \tilde{\boldsymbol{k}}_t^{\mathtt{KR,T}} \\ \mathbf{0} & \mathbf{I}_{10\times 10} \end{bmatrix}, \phi^{\mathtt{AR,T}} \right)
$$

$$
\mathbf{C}_t^{\mathtt{s}} = \begin{bmatrix} 1 & 0 & 0 & 1 & \mathbf{0}_{1\times 10} & 1 & 0 & \beta^{\mathtt{KR,I|T}} & \mathbf{0}_{1\times 10} & \beta^{\mathtt{AR,I|T}} \\ 0 & 0 & 0 & 0 & \mathbf{0}_{1\times 10} & 0 & 1 & 1 & \mathbf{0}_{1\times 10} & 1 \end{bmatrix}
$$

$$
\mathbf{R}_t^{\mathtt{s}} = \text{block diag} \left((\sigma_v^{\mathtt{I}})^2, (\sigma_v^{\mathtt{T}})^2 \right)
$$

$$
\mathbf{Q}_t^{\mathtt{s(s)}} = \text{block diag} \left((\sigma_w^{\mathtt{LT}})^2 \begin{bmatrix} \dfrac{\Delta t^3}{3} & \dfrac{\Delta t^2}{2} & 0 \\ \dfrac{\Delta t^2}{2} & \Delta t & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} \left(\sigma_{w,0}^{\mathtt{KR,I}}\right)^2 & \mathbf{0} \\ \mathbf{0} & \left(\sigma_{w,1}^{\mathtt{KR,I}}\right)^2 \cdot \mathbf{I}_{10\times 10} \end{bmatrix}, (\sigma^{\mathtt{AR,I}})^2, \right.
$$
$$
\left. (\sigma_w^{\mathtt{LL}})^2, \begin{bmatrix} \left(\sigma_{w,0}^{\mathtt{KR,T}}\right)^2 & \mathbf{0} \\ \mathbf{0} & \left(\sigma_{w,1}^{\mathtt{KR,T}}\right)^2 \cdot \mathbf{I}_{10\times 10} \end{bmatrix}, (\sigma^{\mathtt{AR,T}})^2 \right)
$$

$$
\mathbf{Q}_t^{\mathtt{ns(s)}} = \text{block diag} \left(\begin{bmatrix} (\sigma_w^{\mathtt{LT}})^2 \cdot \dfrac{\Delta t^3}{3} & 0 & 0 \\ 0 & (\sigma_w^{\mathtt{LTT}})^2 \cdot \Delta t & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} \left(\sigma_{w,0}^{\mathtt{KR,I}}\right)^2 & \mathbf{0} \\ \mathbf{0} & \left(\sigma_{w,1}^{\mathtt{KR,I}}\right)^2 \cdot \mathbf{I}_{10\times 10} \end{bmatrix}, (\sigma^{\mathtt{AR,I}})^2, \right.
$$
$$
\left. (\sigma_w^{\mathtt{LL}})^2, \begin{bmatrix} \left(\sigma_{w,0}^{\mathtt{KR,T}}\right)^2 & \mathbf{0} \\ \mathbf{0} & \left(\sigma_{w,1}^{\mathtt{KR,T}}\right)^2 \cdot \mathbf{I}_{10\times 10} \end{bmatrix}, (\sigma^{\mathtt{AR,T}})^2 \right)
$$

Non-Stationary Model Matrices $\mathcal{M}^{\text{ns,obs}}$

$$
\mathbf{A}_t^{\text{ns}} \quad = \quad \text{block diag} \left(\begin{bmatrix} 1 & \Delta t & \dfrac{\Delta t^2}{2} \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & \tilde{\boldsymbol{k}}_t^{\text{KR,I}} \\ \mathbf{0} & \mathbf{I}_{10 \times 10} \end{bmatrix}, \phi^{\text{AR,I}}, 1, \begin{bmatrix} 0 & \tilde{\boldsymbol{k}}_t^{\text{KR,T}} \\ \mathbf{0} & \mathbf{I}_{10 \times 10} \end{bmatrix}, \phi^{\text{AR,T}} \right)
$$

$$
\mathbf{C}_t^{\text{ns}} \quad = \quad \begin{bmatrix} 1 & 0 & 0 & 1 & \mathbf{0}_{1 \times 10} & 1 & 0 & \beta^{\text{KR,I|T}} & \mathbf{0}_{1 \times 10} & \beta^{\text{AR,I|T}} \\ 0 & 0 & 0 & 0 & \mathbf{0}_{1 \times 10} & 0 & 1 & 1 & \mathbf{0}_{1 \times 10} & 1 \end{bmatrix}
$$

$$
\mathbf{R}_t^{\text{ns}} \quad = \quad \text{block diag} \left((\sigma_v^{\text{I}})^2, (\sigma_v^{\text{T}})^2 \right)
$$

$$
\mathbf{Q}_t^{\text{s(ns)}} \quad = \quad \text{block diag} \left(\begin{bmatrix} (\sigma_w^{\text{LA}})^2 \cdot \dfrac{\Delta t^5}{20} & 0 & 0 \\ 0 & (\sigma_w^{\text{LTT}})^2 \cdot \dfrac{\Delta t^3}{3} & 0 \\ 0 & 0 & (\sigma_w^{\text{LA}})^2 \cdot \Delta t \end{bmatrix}, \begin{bmatrix} \left(\sigma_{w,0}^{\text{KR,I}}\right)^2 & \mathbf{0} \\ \mathbf{0} & \left(\sigma_{w,1}^{\text{KR,I}}\right)^2 \cdot \mathbf{I}_{10 \times 10} \end{bmatrix}, \right.
$$
$$
\left. (\sigma^{\text{AR,I}})^2, (\sigma_w^{\text{LL}})^2, \begin{bmatrix} \left(\sigma_{w,0}^{\text{KR,T}}\right)^2 & \mathbf{0} \\ \mathbf{0} & \left(\sigma_{w,1}^{\text{KR,T}}\right)^2 \cdot \mathbf{I}_{10 \times 10} \end{bmatrix}, (\sigma^{\text{AR,T}})^2 \right),
$$

$$
\mathbf{Q}_t^{\text{ns(ns)}} \quad = \quad \text{block diag} \left((\sigma_w^{\text{LA}})^2 \begin{bmatrix} \dfrac{\Delta t^5}{20} & \dfrac{\Delta t^4}{8} & \dfrac{\Delta t^3}{6} \\ \dfrac{\Delta t^4}{8} & \dfrac{\Delta t^3}{3} & \dfrac{\Delta t^2}{2} \\ \dfrac{\Delta t^3}{6} & \dfrac{\Delta t^2}{2} & \Delta t \end{bmatrix}, \begin{bmatrix} \left(\sigma_{w,0}^{\text{KR,I}}\right)^2 & \mathbf{0} \\ \mathbf{0} & \left(\sigma_{w,1}^{\text{KR,I}}\right)^2 \cdot \mathbf{I}_{10 \times 10} \end{bmatrix}, (\sigma^{\text{AR,I}})^2, \right.
$$
$$
\left. (\sigma_w^{\text{LL}})^2, \begin{bmatrix} \left(\sigma_{w,0}^{\text{KR,T}}\right)^2 & \mathbf{0} \\ \mathbf{0} & \left(\sigma_{w,1}^{\text{KR,T}}\right)^2 \cdot \mathbf{I}_{10 \times 10} \end{bmatrix}, (\sigma^{\text{AR,T}})^2 \right)
$$

where the superscript I indicates that the matrices belong to the inclinometers, the superscript T indicates that the matrices belong to the thermometers, and $\tilde{\boldsymbol{k}}_t^{\text{KR}} = [\tilde{k}_{t,1}^{\text{KR}}, \tilde{k}_{t,2}^{\text{KR}}, \cdots, \tilde{k}_{t,10}^{\text{KR}}]$ is the normalized kernel values and Δt is the time step at the time t.

Model Matrices for the Extensometers

The model matrices $\{\mathbf{A}_t, \mathbf{C}_t, \mathbf{Q}_t, \mathbf{R}_t\}$ for the stationary and non-stationary model matrices for the extensometers E are defined following

Stationary Model Matrices $\mathcal{M}^{\mathrm{s,obs}}$

$$
\mathbf{A}_t^{\mathrm{s}} = \operatorname{block\,diag}\left(\begin{bmatrix} 1 & \Delta t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \phi^{\mathrm{AR,E}}, 1, \begin{bmatrix} 0 & \tilde{\boldsymbol{k}}_t^{\mathrm{KR,T}} \\ \mathbf{0} & \mathbf{I}_{10\times10} \end{bmatrix}, \phi^{\mathrm{AR,T}}\right)
$$

$$
\mathbf{C}_t^{\mathrm{s}} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & \beta^{\mathrm{KR,E|T}} & \mathbf{0}_{1\times10} & \beta^{\mathrm{AR,E|T}} \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & \mathbf{0}_{1\times10} & 1 \end{bmatrix}
$$

$$
\mathbf{R}_t^{\mathrm{s}} = \operatorname{block\,diag}\left(\left(\sigma_v^{\mathrm{E}}\right)^2, \left(\sigma_v^{\mathrm{T}}\right)^2\right)
$$

$$
\mathbf{Q}_t^{\mathrm{s(s)}} = \operatorname{block\,diag}\left((\sigma_w^{\mathrm{LT}})^2 \begin{bmatrix} \dfrac{\Delta t^3}{3} & \dfrac{\Delta t^2}{2} & 0 \\ \dfrac{\Delta t^2}{2} & \Delta t & 0 \\ 0 & 0 & 0 \end{bmatrix}, (\sigma^{\mathrm{AR,E}})^2, \right.
$$
$$
\left. (\sigma_w^{\mathrm{LL}})^2, \begin{bmatrix} \left(\sigma_{w,0}^{\mathrm{KR,T}}\right)^2 & \mathbf{0} \\ \mathbf{0} & \left(\sigma_{w,1}^{\mathrm{KR,T}}\right)^2 \cdot \mathbf{I}_{10\times10} \end{bmatrix}, (\sigma^{\mathrm{AR,T}})^2\right)
$$

$$
\mathbf{Q}_t^{\mathrm{ns(s)}} = \operatorname{block\,diag}\left(\begin{bmatrix} (\sigma_w^{\mathrm{LT}})^2 \cdot \dfrac{\Delta t^3}{3} & 0 & 0 \\ 0 & (\sigma_w^{\mathrm{LTT}})^2 \cdot \Delta t & 0 \\ 0 & 0 & 0 \end{bmatrix}, (\sigma^{\mathrm{AR,E}})^2, \right.
$$
$$
\left. (\sigma_w^{\mathrm{LL}})^2, \begin{bmatrix} \left(\sigma_{w,0}^{\mathrm{KR,T}}\right)^2 & \mathbf{0} \\ \mathbf{0} & \left(\sigma_{w,1}^{\mathrm{KR,T}}\right)^2 \cdot \mathbf{I}_{10\times10} \end{bmatrix}, (\sigma^{\mathrm{AR,T}})^2\right)
$$

Non-Stationary Model Matrices $\mathcal{M}^{\text{ns,obs}}$

$$\mathbf{A}_t^{\text{ns}} \quad = \quad \text{block diag} \left(\begin{bmatrix} 1 & \Delta t & \dfrac{\Delta t^2}{2} \\[2mm] 0 & 1 & \Delta t \\[2mm] 0 & 0 & 1 \end{bmatrix}, \phi^{\text{AR,E}}, 1, \begin{bmatrix} 0 & \tilde{\boldsymbol{k}}_t^{\text{KR,T}} \\[2mm] \mathbf{0} & \mathbf{I}_{10\times10} \end{bmatrix}, \phi^{\text{AR,T}} \right)$$

$$\mathbf{C}_t^{\text{ns}} \quad = \quad \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & \beta^{\text{KR,E|T}} & \mathbf{0}_{1\times10} & \beta^{\text{AR,E|T}} \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & \mathbf{0}_{1\times10} & 1 \end{bmatrix}$$

$$\mathbf{R}_t^{\text{ns}} \quad = \quad \text{block diag} \left(\left(\sigma_v^{\text{E}}\right)^2, \left(\sigma_v^{\text{T}}\right)^2 \right)$$

$$\mathbf{Q}_t^{\text{s(ns)}} \quad = \quad \text{block diag} \left(\begin{bmatrix} \left(\sigma_w^{\text{LA}}\right)^2 \cdot \dfrac{\Delta t^5}{20} & 0 & 0 \\[3mm] 0 & \left(\sigma_w^{\text{LTT}}\right)^2 \cdot \dfrac{\Delta t^3}{3} & 0 \\[3mm] 0 & 0 & \left(\sigma_w^{\text{LA}}\right)^2 \cdot \Delta t \end{bmatrix}, \right.$$
$$\left. \left(\sigma^{\text{AR,E}}\right)^2, \left(\sigma_w^{\text{LL}}\right)^2, \begin{bmatrix} \left(\sigma_{w,0}^{\text{KR,T}}\right)^2 & \mathbf{0} \\[3mm] \mathbf{0} & \left(\sigma_{w,1}^{\text{KR,T}}\right)^2 \cdot \mathbf{I}_{10\times10} \end{bmatrix}, \left(\sigma^{\text{AR,T}}\right)^2 \right)$$

$$\mathbf{Q}_t^{\text{ns(ns)}} \quad = \quad \text{block diag} \left(\left(\sigma_w^{\text{LA}}\right)^2 \begin{bmatrix} \dfrac{\Delta t^5}{20} & \dfrac{\Delta t^4}{8} & \dfrac{\Delta t^3}{6} \\[3mm] \dfrac{\Delta t^4}{8} & \dfrac{\Delta t^3}{3} & \dfrac{\Delta t^2}{2} \\[3mm] \dfrac{\Delta t^3}{6} & \dfrac{\Delta t^2}{2} & \Delta t \end{bmatrix}, \left(\sigma^{\text{AR,E}}\right)^2, \right.$$
$$\left. \left(\sigma_w^{\text{LL}}\right)^2, \begin{bmatrix} \left(\sigma_{w,0}^{\text{KR,T}}\right)^2 & \mathbf{0} \\[3mm] \mathbf{0} & \left(\sigma_{w,1}^{\text{KR,T}}\right)^2 \cdot \mathbf{I}_{10\times10} \end{bmatrix}, \left(\sigma^{\text{AR,T}}\right)^2 \right)$$

where the superscript E indicates that the matrices belong to the extensometers, the superscript T indicates that the matrices belong to the thermometers, and $\tilde{\boldsymbol{k}}_t^{\text{KR}} = [\tilde{k}_{t,1}^{\text{KR}}, \tilde{k}_{t,2}^{\text{KR}}, \cdots, \tilde{k}_{t,10}^{\text{KR}}]$ is the normalized kernel values and Δt is the time step at the time t.

APPENDIX B   REWARD VALUES

This section provides the grid search procedure in order to obtain the target annual false positive ratios ζ_{FP} of 0.05 and 0.1 per year for the reinforcement learning-based agents. In the process of training the reinforcement learning agents, the only reward value changed is associated with the false positive one. The remaining reward values are fixed and the corresponding values are presented in Table 4.6 in Chapter 4. Table B.1 shows the results of the grid search procedure. The first column is associated with the false positive reward values r^{FP} for training the agents corresponding to the inclinometers. The following three columns show the annual false positive ratios ζ_{FP} obtained. The first column corresponds to the false positive reward values r^{FP} for training the agents associated with the extensometers. The last column shows the resulting annual false positive ratios ζ_{FP}. The sensor simplified names are $\mathtt{I_L}$: Lateral inclinometer, $\mathtt{I_1}$: Longitudinal inclinometer 1, $\mathtt{I_3}$: Longitudinal inclinometer 3, and $\mathtt{E}$: Extensometer. The values in red indicate the annual false positive ratios selected in this thesis.

Table B.1 False positive reward values r^{FP} and corresponding annual false positive ratios ζ_{FP} for inclinometers and extensometers. The sensor simplified names are $\mathtt{I_L}$: Lateral inclinometer, $\mathtt{I_1}$: Longitudinal inclinometer 1, $\mathtt{I_3}$: Longitudinal inclinometer 3, and $\mathtt{E}$: Extensometer.

Reward, r^{FP}	ζ_{FP}			Reward, r^{FP}	ζ_{FP}
	$\mathtt{I_L}$	$\mathtt{I_1}$	$\mathtt{I_3}$		$\mathtt{E}$
-27	0.0322			-140	0.0177
-26	0.0323			-120	0.0177
-25	0.0330			-100	0.0230
-24	0.0440			-80	0.0373
-23	0.0477			-60	0.0440
-22	0.0477			-59	0.0527
-21	0.0478			-58	0.0525
-20	0.0427			-57	0.0563
-19	0.0533			-40	0.0635
-18	0.0548			-30	0.0786
-17	0.0762	0.0404	0.0703	-26	0.0988
-16	0.0670	0.0391	0.0504	-25	0.0846
-15	0.0950	0.0468	0.0614	-20	0.1132
-14	0.1085	0.0647	0.0642	-10	0.3062
-13	0.1342	0.0615	0.0769	-5	1.0084
-12	0.1253	0.0737	0.0840		
-11	0.1318	0.0831	0.0861		
-10	0.1817	0.1086	0.0962		
-9	0.1946	0.0968	0.0948		
-8	0.2397	0.1311	0.1129		

APPENDIX C   SIMULATED STRUCTURAL DAMAGE SCENARIOS

This section provides the changes in structural responses, i.e, $\Delta\theta$ and $\Delta\ell$ due to different damage scenarios, and associated with sensors installed on the Jacques Cartier Bridge (JCB). In Tables C.1-C.5, the damage scenario abbreviated names consist in two or three parts separated by dashes. They represent the damage's type, setup, and severity such that `COR`: corrosion, `BDL`: bearing device lockage, `PST`: partial settlement, `U`: uniform, and `NU`: non-uniform. The sensor simplified names are I_L: Lateral inclinometer, I_1: Longitudinal inclinometer 1, I_3: Longitudinal inclinometer 3, and `E`: Extensometer. The values in red indicate the minimum severity detected by the corresponding RL agent according to its minimum detectable anomaly magnitude $|h_a|_{\min}$.

Table C.1 Simulated structural-damage scenario for the corrosion. The sensor simplified names are I_L: Lateral inclinometer, I_1: Longitudinal inclinometer 1, I_3: Longitudinal inclinometer 3, and `E`: Extensometer.

	Corrosion		$\Delta\theta$, °d			$\Delta\ell$, mm
ID	Severity, %	Abbreviation	I_L	I_1	I_3	E
1	0.5	$COR - 0.5\%$	+0.0000	+0.0003	+0.0001	+0.0031
2	1	$COR - 1\%$	+0.0001	+0.0006	+0.0003	+0.0062
3	2	$COR - 2\%$	+0.0001	+0.0013	+0.0005	+0.0125
4	2.5	$COR - 2.5\%$	+0.0002	+0.0016	+0.0006	+0.0157
5	3	$COR - 3\%$	+0.0002	+0.0019	+0.0007	+0.0189
6	3.5	$COR - 3.5\%$	+0.0002	+0.0023	+0.0009	+0.0221
7	5	$COR - 5\%$	+0.0003	+0.0033	+0.0013	+0.0321
8	5.5	$COR - 5.5\%$	+0.0003	+0.0037	+0.0014	+0.0355
9	6	$COR - 6\%$	+0.0004	+0.0041	+0.0015	+0.0389
10	8	$COR - 8\%$	+0.0005	+0.0056	+0.0021	+0.0529
11	10	$COR - 10\%$	+0.0007	+0.0071	+0.0027	+0.0675
12	12	$COR - 12\%$	+0.0009	+0.0087	+0.0033	+0.0828
13	15	$COR - 15\%$	+0.0011	+0.0114	+0.0042	+0.1070
14	20	$COR - 20\%$	+0.0016	+0.0162	+0.0061	+0.1510
		$|h_a|_{\min}$	0.0020	0.0013	0.0010	0.0090

Table C.2 Simulated structural-damage scenario for the bearing device lockage in the uniform setup.The sensor simplified names are I_L: Lateral inclinometer, I_1: Longitudinal inclinometer 1, I_3: Longitudinal inclinometer 3, and E: Extensometer.

Bearing Device Lockage - Uniform			$\Delta\theta$, °d			$\Delta\ell$, mm
ID	Severity, %	Abbreviation	I_L	I_1	I_3	E
1	5.4	$\text{BDL}-\text{U}-5.4\%$	$+0.0000$	-0.0026	-0.0004	-0.0244
2	10.2	$\text{BDL}-\text{U}-10.2\%$	$+0.0001$	-0.0050	-0.0008	-0.0462
3	14.6	$\text{BDL}-\text{U}-14.6\%$	$+0.0001$	-0.0072	-0.0012	-0.0660
4	22.2	$\text{BDL}-\text{U}-22.2\%$	$+0.0002$	-0.0109	-0.0018	-0.1002
5	36.3	$\text{BDL}-\text{U}-36.3\%$	$+0.0002$	-0.0179	-0.0030	-0.1642
6	53.3	$\text{BDL}-\text{U}-53.3\%$	$+0.0003$	-0.0263	-0.0045	-0.2412
7	74.1	$\text{BDL}-\text{U}-74.1\%$	$+0.0003$	-0.0367	-0.0062	-0.3358
8	85.2	$\text{BDL}-\text{U}-85.2\%$	$+0.0003$	-0.0423	-0.0072	-0.3863
9	92	$\text{BDL}-\text{U}-92\%$	$+0.0003$	-0.0459	-0.0079	-0.4179
10	96.7	$\text{BDL}-\text{U}-96.7\%$	$+0.0003$	-0.0483	-0.0083	-0.4394
11	98.4	$\text{BDL}-\text{U}-98.4\%$	$+0.0003$	-0.0491	-0.0084	-0.4471
12	99.7	$\text{BDL}-\text{U}-99.7\%$	$+0.0003$	-0.0498	-0.0085	-0.4534
13	99.9	$\text{BDL}-\text{U}-99.9\%$	$+0.0003$	-0.0499	-0.0085	-0.4542
14	100	$\text{BDL}-\text{U}-100\%$	$+0.0003$	-0.0500	-0.0085	-0.4546
		$|h_a|_\text{min}$	0.0020	0.0013	0.0010	0.0090

Table C.3 Simulated structural-damage scenario for the bearing device lockage in the non-uniform setup.The sensor simplified names are I_L: Lateral inclinometer, I_1: Longitudinal inclinometer 1, I_3: Longitudinal inclinometer 3, and E: Extensometer.

Bearing Device Lockage - Non-uniform			$\Delta\theta$, °d			$\Delta\ell$, mm
ID	Severity, %	Abbreviation	I_L	I_1	I_3	E
1	5.4	$\text{BDL}-\text{NU}-5.4\%$	$+0.0007$	$+0.0001$	$+0.0002$	-0.0070
2	10.2	$\text{BDL}-\text{NU}-10.2\%$	$+0.0013$	$+0.0002$	$+0.0003$	-0.0135
3	14.6	$\text{BDL}-\text{NU}-14.6\%$	$+0.0019$	$+0.0003$	$+0.0005$	-0.0194
4	22.2	$\text{BDL}-\text{NU}-22.2\%$	$+0.0030$	$+0.0005$	$+0.0007$	-0.0301
5	36.3	$\text{BDL}-\text{NU}-36.3\%$	$+0.0051$	$+0.0007$	$+0.0012$	-0.0512
6	53.3	$\text{BDL}-\text{NU}-53.3\%$	$+0.0079$	$+0.0011$	$+0.0018$	-0.0788
7	74.1	$\text{BDL}-\text{NU}-74.1\%$	$+0.0116$	$+0.0017$	$+0.0026$	-0.1164
8	85.2	$\text{BDL}-\text{NU}-85.2\%$	$+0.0138$	$+0.0021$	$+0.0032$	-0.1385
9	92	$\text{BDL}-\text{NU}-92\%$	$+0.0152$	$+0.0022$	$+0.0035$	-0.1530
10	96.7	$\text{BDL}-\text{NU}-96.7\%$	$+0.0162$	$+0.0024$	$+0.0037$	-0.1633
11	98.4	$\text{BDL}-\text{NU}-98.4\%$	$+0.0166$	$+0.0025$	$+0.0038$	-0.1670
12	99.7	$\text{BDL}-\text{NU}-99.7\%$	$+0.0169$	$+0.0025$	$+0.0039$	-0.1701
13	99.9	$\text{BDL}-\text{NU}-99.9\%$	$+0.0170$	$+0.0025$	$+0.0039$	-0.1705
14	100	$\text{BDL}-\text{NU}-100\%$	$+0.0170$	$+0.0025$	$+0.0039$	-0.1707
		$|h_a|_\text{min}$	0.0020	0.0013	0.0010	0.0090

Table C.4 Simulated structural-damage scenario for partial settlement in the uniform setup. The sensor simplified names are I_L: Lateral inclinometer, I_1: Longitudinal inclinometer 1, I_3: Longitudinal inclinometer 3, and E: Extensometer.

Partial Settlement - Uniform			$\Delta\theta$, °d			$\Delta\ell$, mm
ID	Severity, mm	Abbreviation	I_L	I_1	I_3	E
1	1	$PST - U - 1\,mm$	+0.0000	−0.0013	−0.0013	+0.0000
2	2	$PST - U - 2\,mm$	+0.0000	−0.0026	−0.0025	+0.0000
3	3	$PST - U - 3\,mm$	+0.0000	−0.0038	−0.0038	+0.0000
4	5	$PST - U - 5\,mm$	+0.0000	−0.0064	−0.0064	+0.0000
5	10	$PST - U - 10\,mm$	+0.0000	−0.0128	−0.0128	+0.0000
6	30	$PST - U - 30\,mm$	+0.0000	−0.0384	−0.0384	+0.0004
7	60	$PST - U - 60\,mm$	+0.0000	−0.0768	−0.0768	+0.0017
8	100	$PST - U - 100\,mm$	+0.0000	−0.1279	−0.1279	+0.0048
		$|h_a|_{min}$	0.0020	0.0013	0.0010	0.0090

Table C.5 Simulated structural-damage scenario for the partial settlement in the non-uniform setup. The sensor simplified names are I_L: Lateral inclinometer, I_1: Longitudinal inclinometer 1, I_3: Longitudinal inclinometer 3, and E: Extensometer.

Partial Settlement - Non-uniform			$\Delta\theta$, °d			$\Delta\ell$, mm
ID	Severity, mm	Abbreviation	I_L	I_1	I_3	E
1	1	$PST - NU - 1\,mm$	−0.0030	−0.0010	−0.0009	+0.0021
2	2	$PST - NU - 2\,mm$	−0.0059	−0.0020	−0.0018	+0.0041
3	3	$PST - NU - 3\,mm$	−0.0088	−0.0030	−0.0027	+0.0062
4	5	$PST - NU - 5\,mm$	−0.0147	−0.0050	−0.0045	+0.0103
5	10	$PST - NU - 10\,mm$	−0.0295	−0.0100	−0.0090	+0.0207
6	30	$PST - NU - 30\,mm$	−0.0883	−0.0300	−0.0271	+0.0620
7	60	$PST - NU - 60\,mm$	−0.1766	−0.0599	−0.0543	+0.1240
8	100	$PST - NU - 100\,mm$	−0.2944	−0.0998	−0.0905	+0.2067
		$|h_a|_{min}$	0.0020	0.0013	0.0010	0.0090