Bayesian system identification of civil engineering structures using high resolution optic fibre measurements and surrogate modelling

Andrés Ignacio Martínez Colán

Bayesian system identification of civil engineering structures using high resolution optic fibre measurements and surrogate modelling

by

Andrés Ignacio Martínez Colán

Faculty of Civil Engineering and Geosciences
Delft University of Technology

Supervisors:

Frans van der Meer	TU Delft, chair
Eliz-Mari Lourens	TU Delft
Iuri Rocha	TU Delft
Árpád Rozsas	TNO
Arthur Slobbe	TNO

This report is submitted for the degree of
Master of Science in Civil Engineering

August 2022

Abstract

Bayesian system identification has been extensively adopted in Structural Health Monitoring as a way to probabilistically infer unobservable parameters of the physical model of a structure using measurement data. Combining the Bayesian approach with distributed optic fibre sensors can potentially improve the accuracy and reduce the uncertainty in parameter estimation problems, given the large amount of quasi-continuous data produced by this sensing technology; however, its computational cost could be prohibitively high when using conventional methods since Bayesian inference typically involves a large number of samples, each of which entails a physical model evaluation.

The focus of this work is on performing Bayesian system identification for real-world civil engineering structures within an acceptable running time, while using optic fibre measurements with a high spatial resolution. The proposed methodology employs a cheap-to-compute Gaussian process (GP) surrogate that replaces the main bottleneck of the Bayesian workflow for these type of problems: the evaluation of the log-likelihood. The GP surrogate is actively built by sequentially selecting new training points in areas that are expected to highly contribute to the accuracy of the posterior distribution. Once convergence is achieved, the surrogate is used to obtain the parameter estimates via Markov chain Monte Carlo (MCMC) sampling. Additionally, in order to accelerate the Bayesian workflow, cloud-based parallelization is used to perform multiple finite element analyses simultaneously.

A first synthetic case with an inexpensive frame model is used to test the methodology for problems with two and five probabilistic parameters. An encouraging outcome is obtained with the actively learned GP surrogate, with posterior distributions very close to the full MCMC procedure while requiring a number of physical model evaluations orders of magnitude lower.

After that, a second case study consisting of an existing reinforced concrete bridge with real measurements and a relatively expensive finite element model is investigated. A subset of discrete strain and translation sensors are used to perform an initial parameter estimation that almost exactly resembles the results from previous research on this bridge by Rózsás, et al. (2022), successfully validating the proposed procedure. Then, another parameter estimation is computed using available high resolution optic fibre measurements, after which is shown that the optic fibre provides the best improvements in model predictive capacity among all sensor groups, confirming its potential when combined with Bayesian system identification.

The results of the case studies indicate that the approach presented in this thesis has the capacity to greatly reduce the wall-clock time of Bayesian parameter estimation for real world civil engineering structures with optic fibre measurements, while maintaining a high degree of accuracy. Nevertheless, additional research is required for cases where the statistical parameters governing the measurement and model uncertainty are inferred along with the physical parameters.

Acknowledgements

I would like to thank Arthur Slobbe and Árpád Rozsas, my supervisors from TNO, who patiently introduced me to the world of Bayesian statistics and guided me through this challenging project. Your comments have helped me greatly improve my presentation, writing and programming skills. When I started this thesis, I had no background whatsoever in Python, and thanks to your support, I have written code at a level I never thought I would be able to. For that, I will be forever grateful.

I am thankful to Iuri Rocha for the conversations we had about machine learning and surrogate modelling, and the ideas that were born from them. I also want to express my gratitude to my other TU Delft supervisors Frans van der Meer and Eliz-Mari Lourens for all of their support during the thesis process.

Many thanks to my friends for all the study sessions, as well as the trips and gatherings that kept us sane during the worst parts of the pandemic. Jessica, thank you so much for all the things you have done for me this past year, from the small things, such as bringing food when I was too busy writing, to the big things, like taking care of me for days when I had the accident.

Finalmente, quiero agradecerle a mis padres y hermanas por todo el apoyo que me brindaron durante estos dos últimos años. Poder conversar con ustedes me mantuvo siempre cerca de casa, y aún en mis momentos más difíciles, me sentí respaldado y querido.

Contents

1 Introduction

1.1 Motivation

The continuous and adequate operation of infrastructure is fundamental for sustainable economic growth and societal progress. In developed countries, much of the infrastructure built decades ago requires maintenance, repairment and sometimes replacement. In the Netherlands, the responsible authorities spend approximately €6 billion a year on maintaining the existing infrastructure (TNO, 2014).

Regular manual inspections to assess the state of structures can be costly, on top of requiring specialized labour and equipment. Structural health monitoring (SHM) methods have gained popularity in recent years as they can provide a more effective and consistent way of making decisions related to infrastructure maintenance. In SHM, sensing systems and necessary hardware/software facilities are used to monitor structural responses and operational conditions of structures (Chen, 2018).

A natural way of integrating monitoring data into a physical model in order to obtain more accurate model predictions and/or to reduce the uncertainty in model predictions is provided by Bayesian statistics (Dashti & Stuart, 2015). In this approach, the unknown structural parameters are treated as random variables with a *prior* probability distribution that is then updated into a *posterior* distribution by incorporating the sensing data and its deviation from model predictions, i.e. the *likelihood* of the data. This methodology is usually referred to as Bayesian system identification across SHM literature.

One of the key aspects of SHM is the acquisition of useful measurement data. Standard monitoring practices for civil engineering structures are usually based on a relatively small number of sensing locations using devices such as electric strain sensors, accelerometers, inclinometers, etc. These discrete sensors provide useful data related to local structural behaviour; however, they may miss important information in non-instrumented areas of large structures (Bárrias, et al., 2016). Optic fibre sensors offer an advantage over discrete sensors by providing quasi-continuous strain measurements over their lengths.

Considering the aforementioned points, it is natural to infer that the combination of a robust methodology such as Bayesian system identification with the large amount of data produced by high-resolution optic fibre sensors could potentially improve the predictive performance of structural models; however, solving Bayesian parameter estimation tasks for real civil structures in this context can be computationally challenging for multiple reasons.

First, in most practical applications a closed-form solution of the posterior distribution cannot be obtained, so it has to be numerically approximated using sampling methods. Each of the drawn samples entails a likelihood calculation that requires a finite element analysis, making the procedure unfeasible for models with non-negligible running times, as is usually the case for realistic models in SHM (e.g. with a three-dimensional geometry and a large number of degrees of freedom). In addition to that, the use of a large number of measurements from optic fibre sensors adds another layer of complexity since the model prediction uncertainty of closely spaced points can have a significant degree of correlation. Moreover, the calculation of the likelihood requires obtaining the probability density of a multivariate normal distribution[1], a process that scales cubically with the number of data points, thus, increasing the computational costs even further if optic fibre measurements are used. Considering that research in this context is very limited, this thesis investigates the use of optic fibre measurements for Bayesian system identification of real-world civil engineering structures.

1.2 Research questions

The use of optic fibre sensors for Bayesian system identification of civil engineering structures has great potential, but its computational cost could be prohibitively high when using conventional approaches, which might explain the lack of research on the topic. In view of this, the main research question of the thesis is defined as follows:

How to perform Bayesian system identification for real-world civil engineering structures within an acceptable running time, i.e., in less than 24 hours, while including high-resolution optic fibre measurements?

In order to answer this main research question, the following sub-questions are addressed:

A. *How to include recent developments in surrogate modelling techniques and cloud-based parallelization in order to accelerate the Bayesian inference workflow?*

B. *What gain in information content can be obtained from optic fibre strain sensors compared to conventional discrete sensors for the identification of structural model parameters?*

C. *How to describe the dependency in model uncertainty for optic fibre sensors and what is their influence on the results of Bayesian parameter estimations?*

1.3 Approach

The approach to answer the previously introduced research questions begins with an exploration of the existing literature on the use of optic fibre sensors and/or surrogate modelling for Bayesian inference, in order to identify techniques that might be relevant for this thesis. Then, the methods and tools to be used are selected, including some of the most promising findings of the literature review.

A first case study consisting of a planar frame-type structure with synthetically generated measurements is investigated for 2D and 5D Bayesian parameter estimations. Given its cheap-to-compute structural model, the problem can quickly be solved with and without surrogates, allowing for the proposed methodology to be tested.

Subsequently, a second case study involving a real reinforced concrete bridge in the Netherlands, Bridge 705, is examined. The large amount of available measurement data from discrete sensors and optic fibre allows the study of various sub-cases with different settings,

---

[1] Assuming that Gaussian distributions are used to model the measurement and model prediction uncertainties.

each of which is focused on a particular research question. Also, cloud-based parallelization is used to simultaneously perform multiple finite element analyses of the bridge and reduce the wall-clock time of the Bayesian parameter estimation.

A summary of the case studies is presented on Table 1.

Table 1: Summary of case studies.

Case	Description	Goal	Parallel	Research question	Section
Frame Sub-case 1	2D parameter estimation with synthetic discrete and optic fibre measurements	Test the thesis methodology	No	A	4.2
Frame Sub-case 2	5D parameter estimation using synthetic discrete and optic fibre measurements	Test the thesis methodology	No	A	4.4
Bridge 705 Sub-case 1	2D parameter estimation with real discrete translation and strain measurements	Validate the thesis methodology	Yes	A	6.2
Bridge 705 Sub-case 2	2D parameter estimation with real discrete strain measurements	Compare discrete strain sensors vs optic fibre	Yes	B	6.3
Bridge 705 Sub-case 3	2D parameter estimation with real optic fibre strain measurements	Compare discrete strain sensors vs optic fibre	Yes	B	6.4
Bridge 705 Sub-case 4	5D parameter estimation with real optic fibre strain measurements	Asses influence of uncertainty parameters	Yes	C	6.5

1.4 Scope and limitations

Since the main goal of the thesis is to find a feasible way of performing Bayesian system identification in real world structures while using optic fibre measurements, the emphasis is placed is in prototyping, testing, validating and presenting the methodology, rather than assessing the condition of the structure in the considered real-world case.

For the very same reason, no effort is put on identifying a 'true' probabilistic model class for the data, therefore, the investigation is carried using an assumed model. Furthermore, within the Bayesian framework, the focus is on obtaining posterior distributions, while other aspects such as posterior predictives and model evidences are mentioned but not treated.

In both case studies, only lineal-elastic analyses are performed and temperature loads are not considered. Likewise, research is limited to structures with deterministic static loads, although the proposed approach could theoretically be extended for dynamic loading, granting more research. Even though the only real-world case studied is a concrete bridge, the applicability of the methodology and the conclusions are expected to be valid for civil engineering structures in general.

1.5 Thesis structure

The remaining of the thesis is structured as follows. Chapter 2 contains a literature review related to optic fibre sensors and surrogate modelling in the context of Bayesian system identification. The methods and tools used in this thesis are described in Chapter 3, including the theoretical basis of key parts such as the Bayesian approach and the use of Gaussian process regression for surrogate modelling. Chapter 4 presents the first case study of the thesis, which consists of a planar frame-type structure with synthetic measurements, where two Bayesian parameter estimations are performed with and without surrogates. The second case study, Bridge 705 in the Netherlands, is described in Chapter 5, along with details about the finite element model, the measurement campaign and the settings for the sub-cases; while the results of the parameter estimations are presented in Chapter 6. Finally the answers to the research questions, main thesis contributions and recommendations for future work are treated in Chapter 7.

2 Literature review

Structural system identification generally involves an inverse problem, i.e., the obtention of model parameter values (input) using measurement data of the structure (output). When treated deterministically, inverse problems are typically ill-conditioned and ill-posed, since there is usually not enough measurement information to precisely determine the model (Huang, et al., 2019). In contrast, Bayesian inference provides a probabilistic treatment of the inverse problem that is rational, robust and capable of handling the difficulty of non-unique solutions (Katafygiotis & Beck, 1998). In Bayesian inference, the unknown structural parameters are treated as random variables with prior probability distributions, which are then updated into a posterior distribution by incorporating measurement data via Bayes' theorem.

The Bayesian approach has been extensively applied for various types of civil engineering structures such as buildings, bridges and geotechnical structures; however, the purpose of this chapter is not to provide a comprehensive review of Bayesian inference for structural system identification, for which the reader is referred to Huang et al. (2019), but to highlight the knowledge gap that inspires this thesis and determine relevant methods that might be used to answer the research questions. Consequently, the use of optic fibre sensors and surrogate modelling in the context of Bayesian system identification is be addressed in the next sections.

2.1 Optic fibre sensors

Optic fibre sensors are dielectric devices composed by a fibre core, cladding and coating layers that provide mechanical resistance. The refractive index of the cladding is lower than that of the core, thus the propagation of light is confined to the fibre core only. When the optic fibre suffers a strain or temperature variation, the frequency of the backscattered light is shifted with respect to the reference state, which is then related to a strain value (Bado, et al., 2022).

Optic fibres present multiple characteristics that make them attractive over other sensing technologies, such as their resistance to corrosion and extreme temperatures, their inherent anti-electro-magnetic interference feature and their ease of installation, due to their small diameter, low rigidity and minimal intrusiveness. More importantly, optic fibre sensors are capable of recording continuous strain measurements with monitoring points spaced less than one millimetre, over monitoring lengths varying from centimetres to kilometres (Bado & Casas, 2021).

A typical classification of optic fibres divides them in *quasi-distributed* and *distributed* optic fibre sensors. The former refers to a series of individual sensors, usually *Fibre Bragg Gratings* (FBG), connected through a single optic fibre, that gives discrete strain measurements averaged over the individual gauge lengths. On the other hand, distributed optic fibre sensors provide strain measurements at any point along the fibre, offering a superior spatial resolution.

The use of optic fibre sensors across the SHM literature is extensive and has grown rapidly in recent years, as shown in the reviews by Bárrias, et al. (2016), Harger, et al. (2019), Wu, et al. (2020) and Bado & Casas (2021); however, their application for Bayesian inference is less widespread. This issue is discussed in the rest of the section.

In Zhang & Yang (2013), a damage assessment methodology based on strain modal analysis under Bayesian statistics is proposed and tested on an experimental small steel plate with FBG sensors. An external excitation was applied and the recorded dynamic strains were successfully used to locate the area where damage had been introduced. Another experimental work can be found in Waeytens, et al. (2016) who attached distributed optic fibre sensors to the rebar of a prestressed concrete beam that was statically loaded. The beam model was discretized into several subdomains across its length, each with its own Young's modulus, and Bayesian model updating was used to detect the damaged area, i.e., the subdomain with a diminished Young's modulus. More recently, Schranz, et al. (2022) statically tested a concrete slab strengthened with memory-steel bars and distributed optic fibres attached to them. The recordings were then used to perform Bayesian model updating of a cross-section model based on the output of a Timoshenko beam analysis.

Moving to real civil structures, Zonta, et al. (2013) studied a cable-stayed bridge whose cables had been instrumented with both force sensors and FBG strain sensors. The sources of uncertainty of this case were the parameters relating the actual load intensity on the cable with the recordings, so no finite element model evaluations were required to perform Bayesian parameter estimation. On the other hand, Koune (2021) used a finite element model and optic fibre strain measurements to perform Bayesian system identification on a steel bridge, in order to infer both physical and statistical parameters. In this study, a cheap-to-compute model with Euler-Bernoulli elements was deemed sufficient to accurately describe the structure, thus surrogate modelling was not necessary and the Nested Sampling technique (Skilling, 2006) was directly applied on the actual finite element predictions to sample from the posterior distribution.

To the author knowledge, Febrianto, et al. (2021) is the only example in open literature where optic fibre measurements have been used for Bayesian system identification of a real-world civil engineering structure with a relatively expensive-to-compute finite element model. In this study, the statFEM methodology introduced by Girolami, et al. (2021) is applied to a composite railway bridge instrumented with FBG sensors, in order to obtain model predictions in unobserved locations of the structure, while accounting for different sources of uncertainty. In general, statFEM problems involve the solution of stochastic partial differential equations; however, the formulation of this case allows for the nodal displacement vector to be expressed as a multivariate Gaussian, with a mean equal to the deterministic nodal displacement vector obtained when solving the finite element model for the mean force vector. In other words, the expensive finite element model only has to be evaluated once.

The previous paragraphs have shown that although optic fibres are well-established in SHM, their use on Bayesian inference is limited to a handful of applications that are either experimental or framed in situations that do not require multiple evaluations of a costly physical model. By investigating a real-world structure instrumented with distributed optic fibre sensors, this thesis addresses some of the existing research gap.

2.2 Surrogate modelling

As discussed at the beginning of the chapter, Bayesian statistics provide a powerful framework for solving the inverse problems that many applications of SHM demand; however, for most real cases, Bayesian inference requires a high number of samples, each of which entails a forward model evaluation, making the procedure computationally unfeasible for most realistic civil engineering models. In recent years, surrogate modelling techniques have been introduced into the field to overcome this issue by replacing the forward model (e.g. finite element model) with a cheap-to-compute approximation. Surrogate models are built by strategically selecting samples on which the forward model is evaluated, and then using the input-output pairs to train the approximation function. In the following paragraphs, their use for Bayesian inference across SHM literature is reviewed.

A natural application is to surrogate the individual responses of finite element models that are associated with measurement data using different regression methods. For example, Rózsás, et al. (2022) used a set of grid points to construct a Gaussian process surrogate of static translations and strains in order to perform Bayesian parameter estimation and load identification on a reinforced concrete bridge; while Sbarufatti, et al. (2018) followed a similar approach to probabilistically detect strain-based damage on a synthetic plate but using Artificial Neural Networks. Vibration-based approaches are also present in the literature, such as Ierimonti, et al. (2021), who developed a monitoring scheme that relied on Kriging to surrogate modal frequencies, in order to detect reductions of stiffness in a monumental masonry building; and Pepi, et al. (2019), who used Polynomial chaos surrogates to perform Bayesian system identification on a pedestrian bridge.

It should be noted that the examples introduced beforehand and similar applications are facilitated by low parameter dimensionality, low number of measurements or both. In real-world cases with distributed optic fibre data, building surrogate models of physical responses might still be computationally challenging since the number of physical model evaluations required to be globally accurate, i.e., on the entire parameter domain, could be prohibitively high. On top of that, some surrogating techniques such as Gaussian processes are not inherently prone to multi-output regression, meaning that a very high number of individual response surrogates have to be fit separately, elevating the computing cost.

Strategies based on surrogating the likelihood might alleviate the complications described in the previous paragraph. Indeed, by focusing on the approximation of a single statistical measure rather than multiple physical responses, this approach can result on a simpler formulation of the surrogate, while also allowing for the introduction of active learning algorithms that only select new training points in areas that highly contribute to the accuracy of the desired quantity. Osborne, et al. (2012) proposed a method based on Bayesian quadrature and active learning to approximate the model evidence, i.e., the distribution of observed data marginalized over the parameters, and tested it for a few astrophysical problems. Also in the astrophysics literature, Kandasamy, et al. (2015) developed the *BAPE* method that surrogates the log joint probability density using Gaussian processes and an active learning cost function that combines the uncertainty of the prediction with its mean value. A continuation of this work can be found in Wang & Li (2017), where the surrogated term is the unnormalized posterior divided by a Gaussian mixture approximation of the posterior distribution, with the expectation that this quotient would sequentially give a smoother and flatter function easier to approximate with a Gaussian process. Finally, a somewhat different approach, is given by Acerbi (2018), which combines Variational Bayesian methods with Gaussian process-based Bayesian quadrature to simultaneously deal with the posterior and model evidence.

Multiple solutions have been proposed to minimize the number of likelihood evaluations for problems with expensive-to-compute forward models; however, their applications are mostly encased in astrophysics and neuroscience. The author has not found applications of these strategies for SHM with the exception of Ni, et al. (2021), and Ni, et al. (2022), who respectively used the previously described variational approach and log joint probability surrogate approach, for the Bayesian parameter estimation of a synthetic arch bridge and an experimental small frame structure, i.e., not for a real-world case.

The literature exploration of this section signals a lack of research in the application of surrogating techniques for real-world civil engineering structures where a high number of physical outputs are required, as occurs with high resolution optic fibre sensors. By investigating this scenario, the thesis contributes to reduce the research gap.

3 Methods and tools

In this section, the methods and tools used in this thesis are described. First, Bayes' theorem is introduced, as well as its application to obtain posterior predictive distributions. Then, the data generating process is presented, i.e., the assumed mathematical process that generates measurements from the physical and probabilistic models. Next, the topic of information content of sensors is addressed, as well as methods to evaluate the information gain from prior to posterior distributions. After that, Markov Chain Monte Carlo (MCMC) methods to sample from the posterior distribution are introduced, including specific algorithm such as *Metropolis Hastings* and the *Affine-invariant ensemble sampler*. Additionally, the concepts of surrogate modelling and Gaussian process (GP) surrogates are introduced, along with some GP components, limitations and adaptive sampling strategies for posterior estimation. Finally, the parallelization approach used in this thesis to reduce wall-clock times is presented.

3.1 Bayesian system identification

Many problems in Structural Health Monitoring (SHM) involve the solution of an inverse problem, i.e. inferring the parameters of the physical model of a structure from a set of measurements, rather than obtaining the responses of a structure for a given combination of parameters (forward problem). Dashti & Stuart (2015) explain that solving inverse problems might be difficult due to (a) the presence of measurement noise, for which only its statistical properties are known at best, and (b) the problem being ill-posed, meaning that small changes in the data can induce arbitrarily large changes in the solution.

Adopting the Bayesian approach allows to overcome both of these difficulties. In this way, the solution of the inverse problem is defined as a probability distribution of the parameters with respect to the measurement data and the noise is modelled via its statistical properties. Also, the previous knowledge about the parameters is incorporated as the prior distribution, which can serve as a form of regularization to counteract the ill-posedness (Dashti & Stuart, 2015).

Under the Bayesian approach, the problem is casted as a parameter estimation, often referred to as system identification in the SHM literature. This sub-section presents the main aspects of Bayesian system identification.

3.1.1 Bayes' theorem

Bayesian inference is the process of fitting a statistical model to a set of data and summarizing the result by a probability distribution of the model parameters (Gelman, et al., 2013). The Bayes' theorem for continuous variables reads:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})}{\int p(\boldsymbol{y}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})\, d\boldsymbol{\theta}} \tag{1}$$

where $\boldsymbol{\theta}$ is a vector of parameters defined as uncertain (parameters of interest) and $\boldsymbol{y}$ is a vector of measurements. The knowledge or set of assumptions about $\boldsymbol{\theta}$ is expressed in the *prior* distribution $p(\boldsymbol{\theta})$, which gives a probability density for every possible value of $\boldsymbol{\theta}$ in the parameter domain. The likelihood of the data for a given a parameter combination $\boldsymbol{\theta}$ is given by the *likelihood function* $p(\boldsymbol{y}|\boldsymbol{\theta})$, while $p(\boldsymbol{\theta}|\boldsymbol{y})$ is the *posterior* distribution, i.e., the probability density function of $\boldsymbol{\theta}$ updated with measurements and accounting for the prior. The integral in the denominator $\int p(\boldsymbol{y}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})\, d\boldsymbol{\theta}$ is the *evidence* (also called *marginal likelihood*) of $\boldsymbol{y}$, or $p(\boldsymbol{y})$, and serves as a normalizing term for the numerator so the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ integrates to one over the parameter domain. Although in some situations the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ can be obtained analytically, this is generally not the case and numerical methods are necessary (see Section 3.4).

3.1.2 Posterior predictives

After computing the posterior distribution of theta, it might be desirable to make predictions of observables or variables that depend on these model parameters; however, under the Bayesian approach there is only access to the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ over possible values that the model parameters $\boldsymbol{\theta}$ could take, so this uncertainty needs to be marginalized in order to make predictions (Speagle, 2020):

$$p(\widetilde{\boldsymbol{y}}|\boldsymbol{y}) = \int p(\widetilde{\boldsymbol{y}}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta} \tag{2}$$

where $\widetilde{\boldsymbol{y}}$ is a vector of future observations and $\boldsymbol{y}$ the vector of the already existing observations. The term $p(\widetilde{\boldsymbol{y}}|\boldsymbol{y})$ is known as the *posterior predictive* distribution of $\widetilde{\boldsymbol{y}}$ given $\boldsymbol{y}$, $p(\boldsymbol{\theta}|\boldsymbol{y})$ is the already determined posterior and $p(\widetilde{\boldsymbol{y}}|\boldsymbol{\theta})$ is the likelihood of the future observations. The posterior predictive is the expected value of the likelihood of future observations over the posterior, hence, it incorporates the uncertainty from all sources: model and measurement uncertainty via the likelihood, and parameter uncertainty via the posterior.

3.2 Data generating process

Bayesian data analysis means producing a story of how the data came to be (McElreath, 2016). This process involves the likelihood function introduced in the previous section, and is usually known as the *data generating process*. The content of this section is partially based on the work of Koune (2021).

In this thesis, the data generating process refers to the combination of a physical and a probabilistic model. The physical model is defined as deterministic mathematical representation of a structure for which responses can be obtained as a function of a parameter combination; while the probabilistic model accounts for the intrinsic uncertainties that appear when using a model to describe a physical system. Two sources of uncertainty are considered in this thesis: measurement uncertainty and physical model uncertainty.

Measurement uncertainty refers to the mismatch between the measured quantities and the true responses of the structure caused by sensing errors and/or environmental noise. In this thesis, the measurement error is modelled as an *additive* Gaussian term with zero mean and a standard deviation, which is usually estimated using sensor pre-recordings.

On the other hand, the model uncertainty is related to the discrepancy between the deterministic physical model predictions and the true responses. The errors here mainly obey to mathematical simplifications and idealizations with respect to the real structure. In this thesis, the model prediction error is modelled as a *multiplicative* Gaussian term with a unitary mean. The hypothesis that model uncertainty scales with the model output seems adequate when considering the situation of translation restraints. Indeed, an additive model uncertainty scheme would give unrealistic results here since it is known that displacements in these points are zero by the definition of the problem.

Considering the aforementioned points, the relationship between the measured response $\boldsymbol{y}$ and the deterministic physical model prediction $\boldsymbol{y}_{\text{model}}$ is given by the following equation:

$$\boldsymbol{y} = {\boldsymbol{c}_{\text{model}}}^{T} \cdot \boldsymbol{y}_{\text{model}}(\boldsymbol{\theta}) + \boldsymbol{e}_{\text{meas}} \tag{3}$$

where $\boldsymbol{\theta}$ is the vector of uncertain physical parameters, $\boldsymbol{e}_{\text{meas}}$ is the vector containing the measurement additive noise term and $\boldsymbol{c}_{model}$ is the vector containing the model multiplicative noise term. Considering D as the length of $\boldsymbol{y}$, $\boldsymbol{e}_{\text{meas}}$ is a particular realisation of a random variable $\boldsymbol{E}_{\text{meas}}$ that follows a multivariate Gaussian distribution:

$$\boldsymbol{E}_{\text{meas}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\text{meas}}) \tag{4}$$

$$\boldsymbol{\Sigma}_{\text{meas}} = \sigma_{\text{meas}} \boldsymbol{I}_{(D,D)} \tag{5}$$

where σ_{meas} is the standard deviation of the measurement noise and $\boldsymbol{I}$ is the identity matrix. In a similar way, the model noise $\boldsymbol{c}_{\text{model}}$ is a particular realisation of a random variable $\boldsymbol{C}_{\text{model}}$ with the following distribution (the formulation of the covariance matrix $\boldsymbol{\Sigma}_{\text{model}}$ is discussed later):

$$\boldsymbol{C}_{\text{model}} \sim \mathcal{N}(\boldsymbol{1}, \boldsymbol{\Sigma}_{\text{model}}) \tag{6}$$

Considering $\boldsymbol{M} = \text{diag}(\boldsymbol{y}_{\text{model}}(\boldsymbol{\theta}))$, the covariance matrix $\boldsymbol{\Sigma}_{\text{phys}} = \boldsymbol{M} \cdot \boldsymbol{\Sigma}_{\text{model}} \cdot \boldsymbol{M}$ is defined. Then, the measurements $\boldsymbol{y}$ are realisations of a random variable $\boldsymbol{Y}$ that follows:

$$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{y}_{\text{model}}(\boldsymbol{\theta}), \boldsymbol{\Sigma}_{\text{meas}} + \boldsymbol{\Sigma}_{\text{phys}}) \tag{7}$$

The multivariate Gaussian distribution has a joint probability density given by the following equation (Rasmussen & Williams, 2006):

$$p(\boldsymbol{x}|\boldsymbol{\mu} = \boldsymbol{m}) = (2\pi)^{-D/2} |\boldsymbol{\Sigma}|^{-1/2} \cdot \exp\left(-{}^{1}/_{2}\,(\boldsymbol{x} - \boldsymbol{m})^{T} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{m})\right) \tag{8}$$

where m is the mean vector (of length D), $\boldsymbol{\Sigma}$ is the covariance matrix (of size D, D) and $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$. Since the physical parameter vector $\boldsymbol{\theta}$ only intervenes in the physical model prediction, the likelihood presented in equation (1) can be interpreted as the probability density of having a set of measurements $\boldsymbol{y}$ given a mean of $\boldsymbol{y}_{model}(\boldsymbol{\theta})$:

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{\mu} = \boldsymbol{y}_{\text{model}}(\boldsymbol{\theta})) \tag{9}$$

Defining $\boldsymbol{\Sigma}_{\text{tot}} = \boldsymbol{\Sigma}_{\text{meas}} + \boldsymbol{\Sigma}_{\text{phys}}$, and combining equations (8) and (9) yields the likelihood function:

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = (2\pi)^{-D/2} |\boldsymbol{\Sigma}_{\text{tot}}|^{-1/2} \cdot \exp\left(-{}^{1}/_{2}\,\big(\boldsymbol{y} - \boldsymbol{y}_{\text{model}}(\boldsymbol{\theta})\big)^{T} {\boldsymbol{\Sigma}_{tot}}^{-1} \big(\boldsymbol{y} - \boldsymbol{y}_{\text{model}}(\boldsymbol{\theta})\big)\right) \tag{10}$$

To complete the definition of the data generating process, the covariance matrix $\mathbf{\Sigma}_{\text{model}}$ is addressed. A significant dependence between closely-spaced sensing points is expected, hence, a covariance function of the distance $k(\boldsymbol{x}-\boldsymbol{x}')$ is needed. In this thesis, the *Squared Exponential* (SE) covariance function is used:

$$k(\boldsymbol{x}-\boldsymbol{x}') = \sigma_{\text{model}}^2 \cdot \exp \frac{-(\boldsymbol{x}-\boldsymbol{x}')^2}{l_{\text{corr}}} \tag{11}$$

where $\boldsymbol{x}$, $\boldsymbol{x}'$ are point coordinates in space, while σ_{model} and l_{corr} are the *hyperparameters* of the covariance function, the first representing its amplitude and the second, the length scale or how far the dependence propagates. Many properties of this covariance function make it attractive to use; for example, it is infinitely differentiable, i.e., very smooth, while having only two hyperparameters that need to be tuned. Also, the SE function is stationary, meaning that it only depends on the distance $\boldsymbol{x}-\boldsymbol{x}'$ and not on their absolute positions (Duvenaud, 2014).

Finally, $\mathbf{\Sigma}_{\text{model}}$ is built by evaluating the covariance function k for every pair of sensing points, yielding a square symmetric matrix of size D:

$$\boldsymbol{\Sigma}_{\text{model}\,i,j} = k(\boldsymbol{x}_i - \boldsymbol{x}_j) \tag{12}$$

3.3 Information content of sensors

A visual way of evaluating the information content is by using credible regions. A p-level credible region encompasses p probability mass of the density function (Gelman, et al., 2013), so the credible intervals for each parameter could be plotted for every sub-set of sensors and determine which one is more informative. An illustration of the concept is shown in Figure 1.

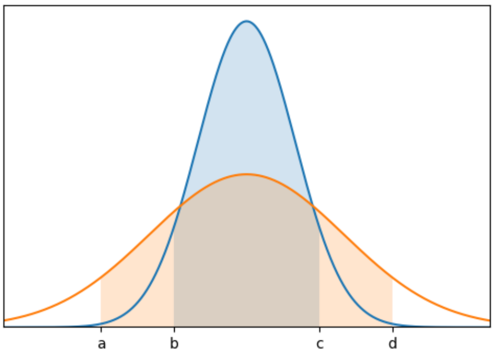


Figure 1: Posterior distributions of a parameter obtained with different sets of sensors. The shadowed regions correspond to the same p-level credible interval. Since the distance between b and c is smaller than the distance between a and d, the information gain of the sensor set associated with the blue curve is higher.

A more principled approach to evaluate the information content is based on the *Kullback-Leibler divergence* which is defined as follows:

$$D(f||g) = \int f(x) \cdot log \frac{f(x)}{g(x)} dx \tag{13}$$

where g and f are probability distributions for a random variable X. The KL divergence can be interpreted as a measure of how far g is from f and it has some of the properties of a distance: for all g, f the divergence $D(g||f) \geq 0$, $D(f||f) = 0$, and conversely if $D(g||f) = 0$, then $g = f$ almost everywhere. However, the KL divergence is not a distance (metric) since it is not symmetric (Commenges, 2015).

The KL divergence of the posterior with respect to the prior could be used to measure the information gain (Huan & Marzouk, 2013):

$$D(p(\boldsymbol{\theta}|\boldsymbol{y})||p(\boldsymbol{\theta})) = \int p(\boldsymbol{\theta}|\boldsymbol{y}) \cdot log \frac{p(\boldsymbol{\theta}|\boldsymbol{y})}{p(\boldsymbol{\theta})} d\boldsymbol{\theta} \tag{14}$$

The idea behind the approach would be to compute the posterior distribution with different sub-sets of sensors separately and compare the values of the KL divergences for each. It is evident that since the KL divergence is obtained by integrating in the whole parameter space, the information gain assessment is global and complements the credible interval approach that gives evaluation on each parameter individually[2].

3.4 Markov Chain Monte Carlo

As mentioned in Section 3.1.1, the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ can usually only be obtained by numerical methods. One approach is to discretize the parameter space in grids and solve the posterior integrals in a Riemann sum-like way; however, the total number of necessary grids increases exponentially with the number of dimensions (*the curse of dimensionality*), making it unattractive even for low-dimensional problems (Speagle, 2020). In this Section, a more appropriate family of computational methods for Bayesian inference is presented: Markov Chain Monte Carlo (MCMC).

In MCMC, a Markov chain[3] of n parameter values is drawn $\{\boldsymbol{\theta}_1 \rightarrow \cdots \rightarrow \boldsymbol{\theta}_n\}$. Considering $m(\boldsymbol{\theta}_\delta)$ the number of iterations from the chain allocated on a parameter region δ, the sample density in that region is (Speagle, 2020):

$$\rho_\delta(\boldsymbol{\theta}) = {}^{m(\boldsymbol{\theta}_\delta)}/_{n} \tag{15}$$

The goal of MCMC methods is to generate a chain (or chains) of parameter values so that after a number of iterations $\rho(\boldsymbol{\theta})$ approximates the posterior $p(\boldsymbol{\theta}|\boldsymbol{y})$. Then, the posterior can be found just by dividing the number of samples on each region by the total number of samples.

A popular approach for MCMC is the *Metropolis-Hastings* (MH) algorithm. Defining the term $\tilde{\mathcal{P}}(\boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})$ as the unnormalized posterior density, a summary of the algorithm is presented:

1. From a position $\boldsymbol{\theta}_i$ draw $\boldsymbol{\theta}'_{i+1}$ from an easy-to-sample proposal distribution $Q(\boldsymbol{\theta}'_{i+1}|\boldsymbol{\theta}_i)$

---

[2] The credible interval approach is global for 1D cases.

[3] A sequence of events where the probability of an event only depends on the previous one.

2. Compute the transition probability[4, 5], i.e., the probability of accepting $\boldsymbol{\theta}'_{i+1}$:

$$T(\boldsymbol{\theta}'_{i+1}|\boldsymbol{\theta}_i) = \min\left[1, \frac{\tilde{\mathcal{P}}(\boldsymbol{\theta}'_{i+1})}{\tilde{\mathcal{P}}(\boldsymbol{\theta}_i)} \frac{\mathcal{Q}(\boldsymbol{\theta}'_{i+1}|\boldsymbol{\theta}_i)}{\mathcal{Q}(\boldsymbol{\theta}_i|\boldsymbol{\theta}'_{i+1})}\right] \tag{16}$$

3. Draw a random number r from a uniform distribution $\sim \mathcal{U}(0,1)$
4. If $r < T(\boldsymbol{\theta}'_{i+1}|\boldsymbol{\theta}_i)$ accept the new point and $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}'_{i+1}$
5. If $r > T(\boldsymbol{\theta}'_{i+1}|\boldsymbol{\theta}_i)$ reject the new point and $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i$

The sequence $\{\boldsymbol{\theta}_1 \to \cdots \to \boldsymbol{\theta}_n\}$ is a Markov Chain since the drawing of $\boldsymbol{\theta}_{i+1}$ was only conditioned on $\boldsymbol{\theta}_i$. The MH algorithm converges to a stationary set of samples from the posterior at $i \to \infty$. Other algorithms can achieve convergence at a faster rate, i.e., with a lower number of likelihood evaluations, and may be preferrable because of that.

One way of measuring the convergence rate is by using the *autocorrelation time*, which can be defined as the amount of iterations necessary for the chain to start independently sampling from the target density (Foreman-Mackey, et al., 2013). Assuming an infinitely long chain $\{\boldsymbol{\theta}_1 \to \cdots\}$ the *auto-covariance* $C(t)$ is defined as:

$$C(t) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}).(\boldsymbol{\theta}_{i+t} - \bar{\boldsymbol{\theta}}) \tag{17}$$

where t is the time lag and $\bar{\boldsymbol{\theta}}$ is the vector of averages. $C(t)$ takes its maximum value when $t = 0$, since the auto-covariance is measured for two identical points. Considering this, the *auto-correlation* $A(t)$ is:

$$A(t) = \frac{C(t)}{C(0)} \tag{18}$$

Using equation (18), the *auto-correlation time* is obtained by:

$$\tau = 2 \sum_{t=1}^{\infty} A(t) \tag{19}$$

As previously mentioned, τ offers a measure of the inverse of the convergence rate. In practice, the generated chains are finite so the auto-correlation time can only be approximated. Many methods to do this are available; however, they are not be described further here.

For this thesis, the MCMC Python package *emcee* (Foreman-Mackey, et al., 2013) is used. This package implements the *affine-invariant ensemble sampler* (Goodman & Weare, 2010), which significantly outperforms standard MH algorithms. Also, this algorithm only requires hand-

---

[4] The transition probability is derived by invoking the principle of Detailed balance which ensures convergence as $i \to \infty$. Under this principle, the probability of moving from $\boldsymbol{\theta}_i \to \boldsymbol{\theta}_{i+1}$ is the same as the probability of $\boldsymbol{\theta}_{i+1} \to \boldsymbol{\theta}_i$ (Speagle, 2020).

[5] This line can be computationally expensive since it involves the evaluation of the likelihood.

tuning of 1 or 2 parameters, compared to $\sim N^2$ for traditional MCMC methods[6], where N is the number of dimensions of the parameter space.

The method, and in particular the so called *stretch move* variant (Goodman & Weare, 2010), comprises an ensemble of K walkers where the proposal distribution for one walker k is determined with respect to the position of the other $K-1$ walkers. The procedure for a single *stretch move* is summarized as follows:

1. Randomly select a walker $\boldsymbol{\theta}_j$ from the complementary ensemble of $K-1$ walkers $\{\boldsymbol{\theta}_1(t+1),\dots,\boldsymbol{\theta}_{k-1}(t+1),\boldsymbol{\theta}_{k+1}(t),\dots,\boldsymbol{\theta}_{K-1}(t)\}$
2. From the distribution $g(z)$, randomly sample a value z
3. Calculate the new proposed position $\boldsymbol{\theta}'_k(t+1)=\boldsymbol{\theta}_j+z(\boldsymbol{\theta}_k(t)-\boldsymbol{\theta}_j)$, see Figure 2
4. Compute the probability of accepting $\boldsymbol{\theta}'_k(t+1)$[7]:

$$T(\boldsymbol{\theta}'_k(t+1)\,|\boldsymbol{\theta}_k(t)\,)=\min\left[1,z^{N-1}\,\frac{\tilde{\mathcal{P}}(\boldsymbol{\theta}'_k(t+1))}{\tilde{\mathcal{P}}(\boldsymbol{\theta}_k(t))}\right] \tag{20}$$

5. Draw a random number r from a uniform distribution $\sim\mathcal{U}(0,1)$
6. If $r<T(\boldsymbol{\theta}'_k(t+1)\,|\boldsymbol{\theta}_k(t)\,)$ accept the new point and $\boldsymbol{\theta}_k(t+1)=\boldsymbol{\theta}'_k(t+1)$
7. If $r>T(\boldsymbol{\theta}'_k(t+1)\,|\boldsymbol{\theta}_k(t)\,)$ reject the new point and $\boldsymbol{\theta}_k(t+1)=\boldsymbol{\theta}_k(t)$

The distribution $g(z)$ proposed by Goodman & Weare (2010) is:

$$g(z)=\begin{cases}1/\sqrt{z} & \text{if z} \in [0.5,2.0]\\ 0 & \text{otherwise}\end{cases} \tag{21}$$

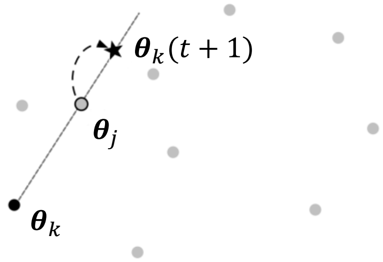


Figure 2: A *stretch move*. The grey dots are the walkers that are not used in this move. The proposal is generated along the straight line connecting $\boldsymbol{\theta}_j$ and $\boldsymbol{\theta}_k$. Adapted from Goodman & Weare (2010).

3.5 Surrogate modelling

3.5.1 General concepts

Many engineering tasks such as model calibration and optimisation require the evaluation of a forward model for a number of times that might be practically unfeasible for heavy simulations (e.g. the three-dimensional finite element model of a large structure). Surrogate

---

[6] A common proposal $Q(\boldsymbol{\theta}'_{i+1}\,|\boldsymbol{\theta}_i)$ is a Gaussian distribution centred on $\boldsymbol{\theta}_i$ with a covariance matrix tuned for performance. This means adjusting $N(N+1)/2$ parameters corresponding to the unique terms in the symmetric covariance matrix.

[7] This line can be computationally expensive since it involves the evaluation of the likelihood.

modelling offers an alternative to alleviate this difficulty. Considering a function that maps a series of inputs to outputs, a *surrogate model* is a cheap-to-compute replacement of the function that uses a set of training data to make predictions of the function output in points where the forward model has not been evaluated.

Surrogate models deal with problems that can be classified as *regression*, when the output to predict is a continuous quantity, or *classification*, when the output is a discrete label. Also, a surrogate is either *global*, if the surrogate is built with the aim of making predictions over the entire parameter space, or *local*, if predictions are only necessary in specific regions of the parameter space (e.g. around maxima for optimisation).

A surrogate model can be constructed using a fixed and previously obtained dataset; however, in problems where the forward model is computationally expensive and there is no training data yet, the number of points where the model is evaluated should be minimized. Here, an *adaptive* sampling technique may be preferable. Under this approach, the surrogate is built in a sequential manner, fitting the function at every step and drawing samples (and evaluating the real model) based on the prediction of the surrogate, which is then used to refit it at the next step. This procedure is followed until a convergence threshold is reached.

Two approaches may be followed in adaptive sampling: *exploitation* and *exploration*. Exploitation refers to adding more samples in regions that are already providing good results (e.g. the lowest output values for a minimisation problem), while exploration focuses on under-sampled areas that may provide even better results. Normally, both are necessary since an exploration step precedes the finding of an interesting area to exploit. On the opposite side, if no exploitation takes place, then the surrogate construction does not need to be adaptive and a scheme that samples the parameter space evenly, such as *Latin Hypercube*, could be employed. The goal of adaptive sampling techniques is to find the right balance between these two strategies.

This thesis only uses *Gaussian process regression* surrogates, which are described in the following section; nevertheless, the reader should be aware that many families of surrogating methods exist, such as *linear regression*, *decision trees*, *support vector machines*, *neural networks*, among others.

3.5.2 Gaussian process regression

Gaussian processes (GP) are a generalization of multivariate Gaussian probability distributions. A formal definition is given by Rasmussen and Williams (2006): *"A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution"*. The GP model of a function $f(\boldsymbol{x})$ is defined by its mean function $m(\boldsymbol{x})$ and covariance function or *kernel* $k(\boldsymbol{x},\boldsymbol{x}')$:

$$m(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})] \tag{22}$$

$$k(\boldsymbol{x},\boldsymbol{x}') = \mathbb{E}\big[\big(f(\boldsymbol{x}) - m(\boldsymbol{x})\big)\big(f(\boldsymbol{x}') - m(\boldsymbol{x}')\big)\big] \tag{23}$$

and the Gaussian process is written as:

$$f(\boldsymbol{x}) \sim \mathcal{GP}\big(f(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\big) \tag{24}$$

GP's have several properties that make them attractive for regression. For example, they offer a high amount of flexibility to approximate different types of functions, and can even be universal function approximators depending on the choice of the *kernel* (Micchelli, et al., 2006). In addition to that, GP's require relatively few parameters to be estimated (e.g. compared to neural networks), lessening the need for complex optimisation (Duvenaud, 2014). Furthermore, a GP predictive distribution can be analytically computed given a *kernel* and some fixed observations.

It is a common practice to assume that the mean function $m(\boldsymbol{x})$ is simply zero everywhere, since the uncertainty about the mean function can be taken into account by adding an extra term to the kernel (Duvenaud, 2014). Figure 3 shows an example with functions drawn from a Gaussian Process:

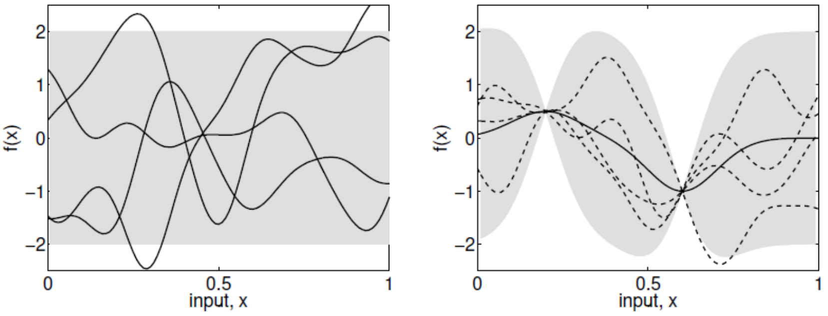


Figure 3: Functions randomly drawn from a GP prior with zero mean and a unitary standard deviation (left) and from the GP posterior after two data points have been observed (right). The grey shaded area corresponds to the limits set by the mean plus and minus two times the standard deviation (Rasmussen & Williams, 2006).

Considering the regression problem $y = f(\boldsymbol{x}) + \varepsilon$, with a Gaussian noise term $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$, the joint distribution of the observed target values y and the predicted values f_*:

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{K}(\boldsymbol{X},\boldsymbol{X}) + \sigma_n^2\boldsymbol{I} & \boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}_*) \\ \boldsymbol{K}(\boldsymbol{X}_*,\boldsymbol{X}) & \boldsymbol{K}(\boldsymbol{X}_*, \boldsymbol{X}_*) \end{bmatrix}\right) \tag{25}$$

where $\boldsymbol{X}$ and $\boldsymbol{X}_*$ contain the training points and prediction points, respectively, while the terms $\boldsymbol{K}$ denote the matrices of the covariances evaluated element-wise on every pair of training and prediction points. The mean and covariance of the predictive distribution f_* are given by:

$$\bar{f}_* = \boldsymbol{K}(\boldsymbol{X}_*,\boldsymbol{X})[\boldsymbol{K}(\boldsymbol{X},\boldsymbol{X}) + \sigma_n^2\boldsymbol{I}]^{-1}y \tag{26}$$

$$\mathrm{cov}(f_*) = \boldsymbol{K}(\boldsymbol{X}_*, \boldsymbol{X}_*) - \boldsymbol{K}(\boldsymbol{X}_*,\boldsymbol{X})[\boldsymbol{K}(\boldsymbol{X},\boldsymbol{X}) + \sigma_n^2\boldsymbol{I}]^{-1}\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}_*) \tag{27}$$

Simplified expressions can be obtained for a single test point x_*. Considering $\boldsymbol{K} = \boldsymbol{K}(X,X)$ and $k_* = \boldsymbol{K}(X,x_*)$:

$$\bar{f}_* = {k_*}^T(\boldsymbol{K} + \sigma_n^2\boldsymbol{I})^{-1}y \tag{28}$$

$$\mathbb{V}[f_*] = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k_*{}^T (\boldsymbol{K} + \sigma_n^2 \boldsymbol{I})^{-1} k_* \tag{29}$$

The kernel $k(\boldsymbol{x}, \boldsymbol{x}')$ of a GP expresses the similarity between two values of a function when it is evaluated in $\boldsymbol{x}$ and $\boldsymbol{x}'$, or in other words, it determines which functions are more likely under the GP prior (Duvenaud, 2014). Since the mean prior function is usually zero, as explained before, all of the prior knowledge of the function $f(\boldsymbol{x})$ is deposited on the kernel. Therefore, the kernel should be able to capture features of the function structure that are previously recognised such as *symmetry*, *periodicity* or *exponential decay*.

Figure 4 and Figure 5 show some kernels commonly used in GP's. Notice that every kernel contains a scaling factor σ_f. The term l denotes the correlation length, or how far it takes for two points to be uncorrelated, while p represents the period or the distance between repetitions, and c is the offset at which the lines of the posterior cross the abscise. The term α determines the relative weighting between small and large scale variations and when $\alpha \to \infty$, the Rational Quadratic kernel turns into a Square Exponential kernel. Notice that δ is the Kronecker-delta function.

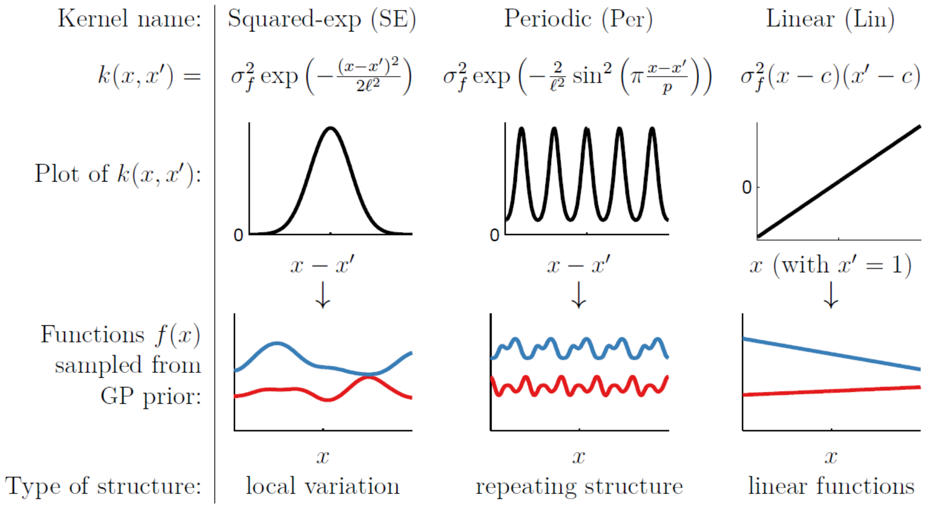


Figure 4: Basic kernels commonly used in GPs – Part 1 (Duvenaud, 2014).

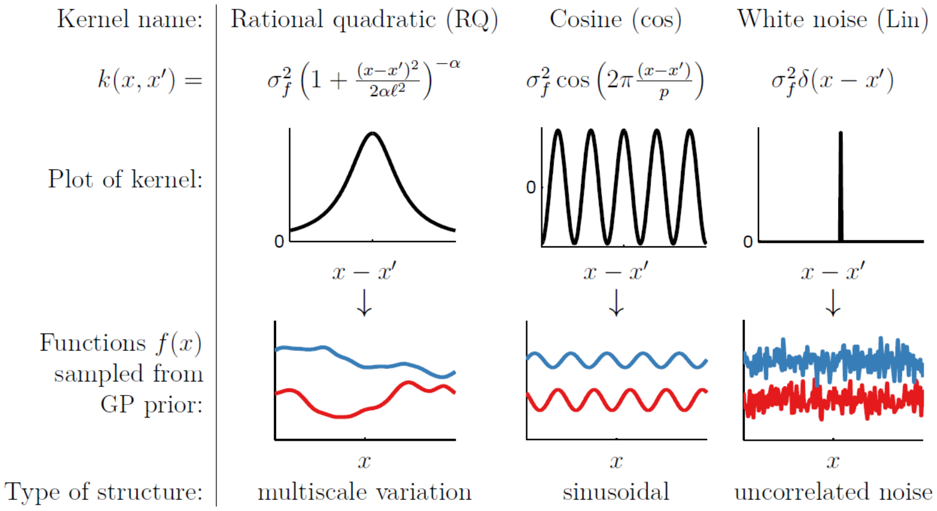


Figure 5: Basic kernels commonly used in GPs – Part 2 (Duvenaud, 2014).

The shown kernels can be summed or multiplied in order to construct new kernels that combine the desired properties, for example, a typical kernel structure sums a Square exponential kernel with a White noise kernel, to have both model correlations and observation noise. Once the structure of the kernel has been chosen, a question arises: *what values should the hyperparameters have?* This issue is of great relevance since the hyperparameters greatly influence the posterior functions.

Rasmussen & Williams (2006) show a procedure to set the values of the hyperparameters by maximizing the *marginal likelihood*[8] with respect to the hyperparameters ψ using partial derivatives:

$$\frac{\partial}{\partial \psi_i} \log p(y|\boldsymbol{X}, \psi) = \frac{1}{2} \operatorname{tr}\left((\alpha\alpha^T - \boldsymbol{K}^{-1}) \frac{\partial \boldsymbol{K}}{\partial \psi_i} \right) \tag{30}$$ [9]

Notice that Equations (28), (29) and (30) involve the computation of the inverse matrix $\boldsymbol{K}^{-1}$, an operation which scales as $\mathcal{O}(n^3)$ where n is the number of rows in the square matrix. Also, the storage of covariance matrices scales as $\mathcal{O}(n^2)$. This means that GP prediction may become prohibitively expensive for training sets consisting of more than a few thousand points, especially when the surrogate construction is expected to be refined step-wise, as in this thesis. Many approximated methods have been proposed for Gaussian processes with a large number of samples, most of them using a carefully chosen subset of training points (Rasmussen & Williams, 2006); however, the details of these methods are not discussed further in this thesis and only exact Gaussian process regression is used.

3.5.3 Adaptive sampling for posterior estimation

A natural way of applying the techniques from this chapter into a civil engineering problem is to use Gaussian processes to surrogate the individual finite element responses (e.g. strains) at positions where sensors are located. Here, an adaptive sampling strategy is preferable over one-time training since it has the potential to minimize the number of forward model evaluations, which is expected to be the bottleneck of the Bayesian system identification workflow of this thesis.

The idea would be to build a *global* GP (see 3.5.1) with an initial training set and then draw new points in areas where the GP predicts the maximum variance. Finding these new points results in an optimisation problem. Convergence could be checked by comparing the maximum variance in the current step with the maximum variance at the beginning.

One way of implementing the mentioned approach is to separately and sequentially fit surrogates for every sensor response, reusing exact function evaluations from the previous surrogates for the new ones, although this would be too cumbersome for the problems analysed in this thesis due to the high number of data points of an optic fibre sensor (e.g., 10 GP surrogates per meter of optic fibre). Also, given that the outputs are correlated, valuable

---

[8] Marginalized over the model parameters $\boldsymbol{\theta}$: $p(y|\boldsymbol{X}, \psi) = \int p(y|\boldsymbol{X}, \boldsymbol{\theta}).p(\boldsymbol{\theta}|\,\psi)d\boldsymbol{\theta}$.

[9] $\frac{\partial K}{\partial \psi_i}$ is an element-wise partial derivative of K over ψ_i.

information may be lost when using separate surrogates. One could also include the location along the optic fibre as an additional parameter; however, every finite element evaluation would produce such a large number of samples (e.g. 10 samples per meter of optic fibre) that the computations would quickly become unmanageable for Gaussian process regression due to its cubic time complexity[10].

Another way of doing this is to build a multiple-output surrogate model, which would appear more promising since the knowledge of the physical model could be introduced to constrain and interpolate outputs, thus reducing the number of forward model evaluations. A multiple output GP surrogate is presented in Equation (31):

$$f(x) \sim \mathcal{GP}\big(0, \boldsymbol{K}_M(\boldsymbol{x}, \boldsymbol{x}')\big) \tag{31}$$

where $\boldsymbol{K}_M(\boldsymbol{x}, \boldsymbol{x}') \in R^{TxT}$ is the multiple-output covariance matrix, with T the number of outputs. The performance of the multiple-output GP depends on $\boldsymbol{K}_M$ being able to capture the output correlations and transferring available information across outputs (Liu, et al., 2018), which is non-trivial and still an active field of research.

An alternative to the previous strategies is to surrogate the likelihood function instead of the individual physical model responses. In this way, the problem of dealing with either multiple single-output GP surrogates or a single multiple-output GP with a complicated output covariance matrix is avoided. Also, an additional computational gain is obtained by predicting the likelihood values from a GP rather than calculating Equation (10), and its expensive inverse matrix operations.

On this line, Kandasamy, et al. (2015) proposed an adaptive sampling strategy for a GP containing the likelihood function. The *Bayesian active posterior estimation* method (BAPE) requires the construction of a GP that surrogates the *log-joint* probability:

$$\log\big(p(\boldsymbol{\theta}, \boldsymbol{y})\big) = \log\big(p(\boldsymbol{\theta}|\boldsymbol{y}) \cdot p(\boldsymbol{\theta})\big) \tag{32}$$

Then, the algorithm selects new samples in areas where the GP predicts the log-joint probability to have the highest uncertainty. In contrast with the approaches mentioned before, the plain standard deviation is not a good predictor of uncertainty for BAPE since the GP is in the log domain. Indeed, areas with a low likelihood would have high negative log-joint probabilities, resulting in high standard deviations. The authors propose to use the *exponentiated variance* as a more adequate measure of uncertainty for the log-joint probability. The Exponentiated variance utility function for a sample $\boldsymbol{\theta}$ is:

$$u_{\mathrm{EV}}(\boldsymbol{\theta}) = exp(2\mu(\boldsymbol{\theta}) + \sigma^2(\boldsymbol{\theta}))\,(exp\,(\sigma^2(\boldsymbol{\theta}) - 1)) \tag{33}$$

---

[10] This approach would be interesting for surrogating techniques that are able to handle large datasets (e.g. neural networks).

where $\mu(\boldsymbol{\theta})$ and $\sigma(\boldsymbol{\theta})$ are the log-joint probability mean and standard deviation predicted by the GP at point $\boldsymbol{\theta}$. A new sample $\boldsymbol{\theta}'$ is chosen at the point where the utility function is maximized:

$$\boldsymbol{\theta}' = \mathrm{argmax}(u_{\mathrm{EV}}(\boldsymbol{\theta})) \tag{34}$$

The utility function u_{EV} reduces high variances in areas with low likelihood and amplifies small variances in areas with high likelihood, which means that the focus is only on regions that highly contribute to the posterior distribution accuracy, rather than on the entire parameter domain. It is expected then that the number of samples required with BAPE is considerably lower than for strategies that use global surrogate for individual physical responses. Because of this reason, BAPE is the surrogating approach used for this thesis.

It should be noted that BAPE also carries some drawbacks. For example, the GP surrogate is not independent of the measurement data since it includes the log likelihood, therefore it cannot be reused for different parameter estimation cases (e.g. when using different subsets of sensors). Moreover, the likelihood can be highly nonlinear and fast-varying, making it a difficult function to approximate with a Gaussian process (Wang & Li, 2017). Additionally, the algorithm is focused on being accurate in areas with high posterior density, which is efficient for the estimation of the posterior distribution, but does not offer a straightforward path to estimate the evidence and posterior predictives.

The final ingredient of the adaptive surrogate is the convergence criterion, i.e., when to stop drawing samples. For this thesis, convergence is checked by comparing the variation of the posterior mean of every parameter at each step, normalized with its posterior standard deviation, as proposed by Fleming & VanderPlas (2018). In this way, convergence is controlled by quantities that are directly related to the physical problem, making it more intuitive than approaches based on more abstract concepts such as the KL divergence. The convergence procedure for an adaptive step t is shown below:

1. Perform MCMC using the log joint probability GP fitted with the samples of the previous $t-1$ adaptive steps
2. Use the MCMC samples to compute the means $\mu_{i,t}$ and standard deviations $\sigma_{i,t}$ of each parameter θ_i
3. For each θ_i, compute the term $\left(\mu_{i,t} - \mu_{i,t-1}\right)/\sigma_{i,t-1}$
4. If the term is below a preestablished threshold ϵ for all θ_i, convergence is reached. If not, the algorithm moves to step $t+1$

3.6 Cloud-based parallelization

Even though the use of surrogating techniques can significantly decrease the total computation time of Bayesian system identification for real-world structures, multiple forward model evaluations are still needed for the different parameter combinations required to train the surrogate model. A strategy to further accelerate the Bayesian workflow is to run the finite element analyses in a parallel way instead of sequentially.

The ability to perform parallel finite element model evaluations in a single computer is limited by its memory and processing power. On the other hand, building a local IT infrastructure capable of handling these parallel tasks would have a high initial cost and may be wasteful during the times where it operates below its capacity. Cloud-based parallelization offers a more efficient way of approaching this problem by allowing to get the desired amount of computational resources on demand, and only during the time the parallel tasks are executed. For this thesis, a Python-based cloud parallelization tool developed by the *Netherlands Organisation for Applied Scientific Research* (TNO) was used. A brief description of how the parallelization tool operates is presented below

First, a vector of model parameter combinations and a base folder containing the finite element analysis model are passed. Then, a task folder is created for each element of the parameter vector and the contents of the base folder are copied into them. After that, the input parameters are used as arguments in a predefined pre-processing function that applies changes to the model files, which are then uploaded to an external cluster to start the cloud-based parallel execution of tasks. Finally, once the parallel finite element model runs are finished, the output analysis files are downloaded to the local machine and a post-processing function is applied to obtain the desired results for each parameter combination.

The cloud service used in this case is provided by *Microsoft Azure*, which among other things, allows for the execution of tasks in *virtual computers* that run in a physical host operated by them. On top of that, *Argo Workflows* and *Kubernetes* are employed, the first one being responsible for the creation and management of parallel workflows, while the second controls the execution and communication with *Azure*. The implementation allows for the user to define the amount of resources to employ for each task, such as the memory and number of virtual CPUs, meaning that the amount of finite element parallel runs is only limited by the number of available software licenses (if required).

4 Case study 1: Synthetic case

In this chapter, some of the methods described previously are applied in order to complete the Bayesian system identification of an in-plane frame-type structure. Specifically, the posterior distributions of the physical parameters are obtained for both a two-dimensional problem (4.1 and 4.2) and a five-dimensional problem (4.3 and 4.4), using sets of synthetic measurements that combine strain fibre-optic sensors and discrete translation sensors. The goal of this case study is to illustrate some of the concepts of the Bayesian system identification workflow and explore certain features that may be computationally infeasible to do for a real world case.

The parameter estimations are performed using the MCMC approach shown in Section 3.4 under two schemes: the first evaluates the log-likelihood for a set of parameters by calling the FE model directly, while the second scheme uses an adaptively-built GP surrogate of the log joint probability (i.e., log prior plus log likelihood) and could be more suitable for problems with expensive-to-compute forward models (see Section 3.5.3).

The structure consists of a continuous three-span rectangular beam with a uniformly distributed load on top and two rectangular columns (see Figure 6). The beam–column connections are assumed to be rigid. The geometry and load parameters are considered as deterministic.

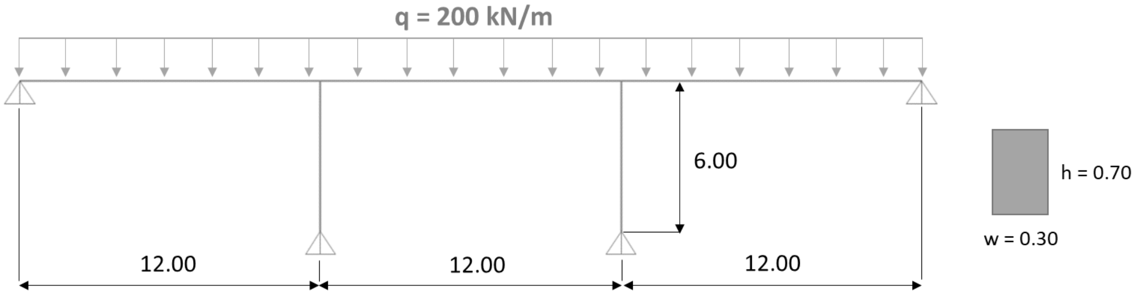


Figure 6: Elevation view of the frame-type structure and cross-section for both columns and beams. Dimensions are in meters.

The finite element model consists of five 2-node Euler-Bernoulli beam-column elements. The beam ends (side supports) are fixed against translations but they can freely rotate. The same applies for the base of the columns; however, in this case vertical settlements are set for both the 2D and 5D cases as parameters to be recovered by the Bayesian system identification.

4.1 Description of the problem – 2-dimensional problem

Figure 7 shows the setup of the 2D case. Two parameters are defined as random variables: the Young's modulus E of the structure and the vertical settlement Δ_1 of the left column base.

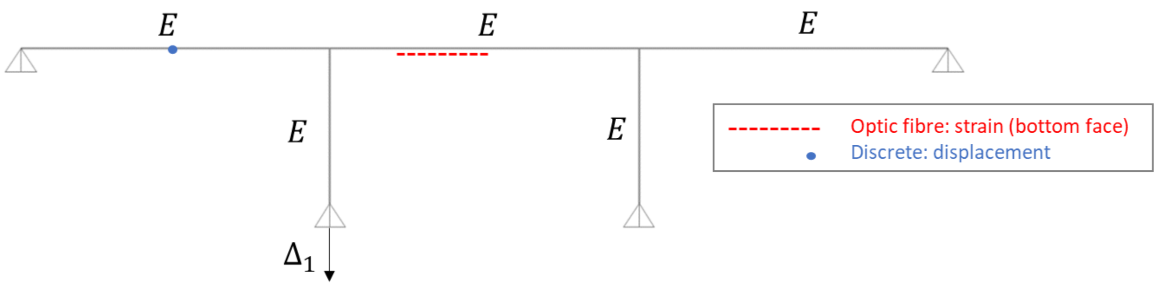


Figure 7: Probabilistic parameters and set of sensors - 2D case.

A discrete vertical displacement sensor is positioned at midspan of the left beam, while a 3 m long, high-resolution fibre-optic strain sensor is placed on the bottom face of the middle beam starting at L/4. The latter sensor is discretised every 120 mm, resulting in 26 data points.

The prediction of the FE model is implemented through a function that takes two parameters as input and returns a vector with 1 displacement and 26 strains in the location of the sensors, as is shown in Equation (35).

$$\boldsymbol{x}_{\text{FEM}}(E,\,\Delta_1) = ((\delta 1_1);(\varepsilon 2_1,\,\varepsilon 2_2,\dots,\varepsilon 2_{26})) \tag{35}$$

The synthetic measurements are produced under the data generating process described in Section 3.2. Figure 8 shows the correlation matrices for the measurement and model uncertainty.

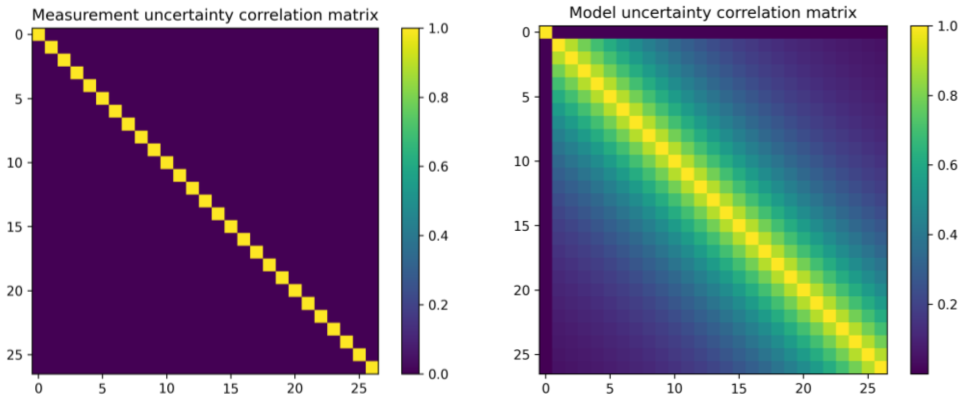


Figure 8: Tile plots of the measurement uncertainty correlation matrix (left) and model uncertainty correlation matrix (right) – 2D case. Each pixel represents the correlation between two measurement data points.

Table 2 shows the values of the physical and probabilistic parameters used. For a successful Bayesian parameter estimation, it is expected that these input values are recovered.

Table 2: Parameters for synthetic measurements generation – 2D case.

Parameter	Value
E_{true}	25 GPa
$\Delta_{1\ \text{true}}$	10 mm
σ_{δ}	0.2 mm
σ_{ε}	0.5 mm
σ_{model}	0.02
l_{corr}	1000 mm

Table 3 and Figure 9 show the prior distribution.

Table 3: Prior distribution – 2D case. The random variables are assumed to be mutually independent.

Parameter	Prior distribution
E	$\mathcal{N}(30, 10)\ GPa$
Δ_1	$\mathcal{N}(15, 5)\ mm$

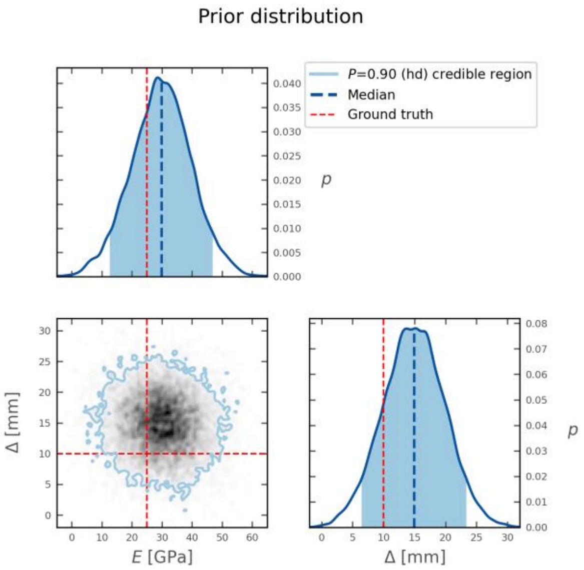


Figure 9: Prior distribution – 2D case. The red and blue dashed lines represent the ground truth values of the parameters and the estimated medians, respectively. Diagonal: 1D marginal densities. Off diagonal: 2D contour plot of joint probability density.

4.2 Results – 2-dimensional problem

In this section, the ground truth parameters for E and Δ_1 are recovered with Bayesian parameter estimation using both a direct MCMC approach and a Gaussian process surrogate approach. The latter is expected to reduce the number of finite element model evaluations, which is typically a bottleneck in Bayesian system identification of real world structures.

4.2.1 Direct MCMC estimation

The MCMC-based parameter estimation of this section is performed with the help of the Python package *emcee* (Foreman-Mackey, et al., 2013) introduced in Section 3.4. The number of walkers is set at 20 and the sampling initial positions are placed at the mean of the prior distribution with small random variations for each walker.

One of the challenges of MCMC methods is to determine how many samples are sufficient/needed. If too few steps are run, the obtained samples are not be independent from the initial positions of the walkers. In contrast, too many steps may be unnecessary to obtain accurate-enough posterior distributions and may be computationally expensive since every sample requires the calculation of the likelihood, which needs the evaluation of the forward model. This might be especially challenging for structures with heavy FE models; however, in this case study the FE model is simple and cheap-to-evaluate, hence we can run a high number of steps at a low computational cost.

Figure 10 shows the trace plots of the two physical parameters for a single walker. It can be seen that they both start near their prior mean and after approximately 100 steps, they start sampling around their ground truth values. The initial phase where the sampling is not independent from the initial positions is called *burn-in phase*. The burn-in phase is discarded when estimating the posterior distributions.

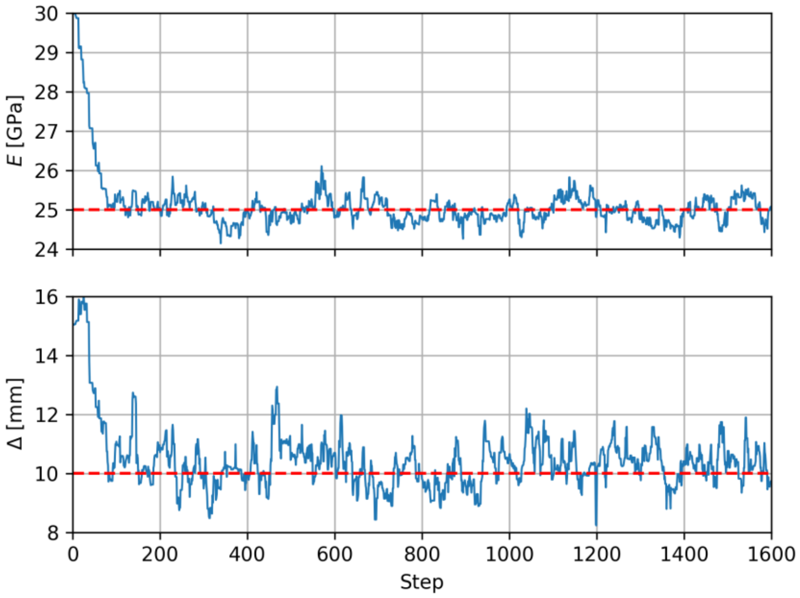


Figure 10: Trace plots for a single walker on 1600 steps – 2D case. The red dashed lines and blue continuous lines represents the ground truth values of the parameters and the samples on every step, respectively.

Table 4 presents a summary of the marginal distribution containing the means, medians and standard deviations, as well as the 25th, 75th and 95th percentiles.

Table 4: Summary of the posterior distribution – 2D case.

Parameter	True val.	Mean	Median	Std. dev.	25th per.	75th per.	95th per.
E (GPa)	25.00	25.01	25.00	0.30	24.80	25.21	25.51
Δ_1 (mm)	10.00	10.12	10.12	0.74	9.62	10.60	11.35

It is clear from Figure 11 that there is a considerable reduction of the uncertainty from the posterior with respect to the prior shown in Figure 9. There is also a shift towards the ground truth values in both parameters. These results show that the Bayesian parameter estimation is successful. Additionally, a positive correlation of 0.31 is present between these parameters.

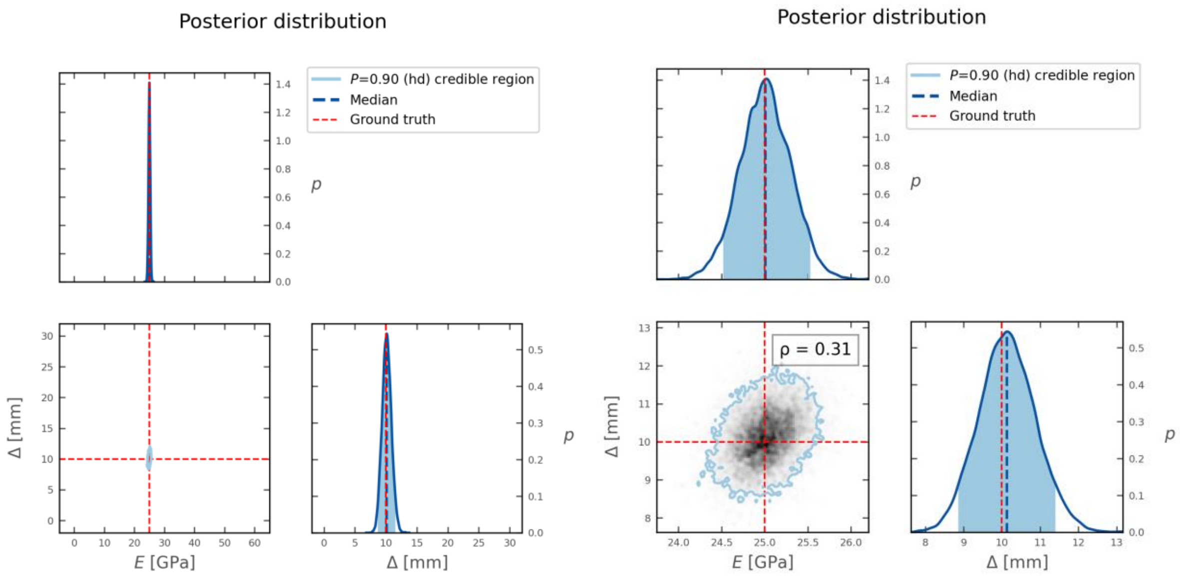


Figure 11: Posterior distribution – 2D case. The left plot shows the same axis limits as the prior while the right one presents a zoomed-in view. The red and blue dashed lines represent the ground truth values of the parameters and the estimated medians, respectively. Diagonal: 1D marginal densities. Off diagonal: 2D contour plot of joint probability density.

4.2.2 BAPE surrogate model

In this section, the parameter estimation is performed with the adaptive sampling strategy described in Section 3.5.3, using the Python package *approxposterior* (Fleming & VanderPlas, 2018).

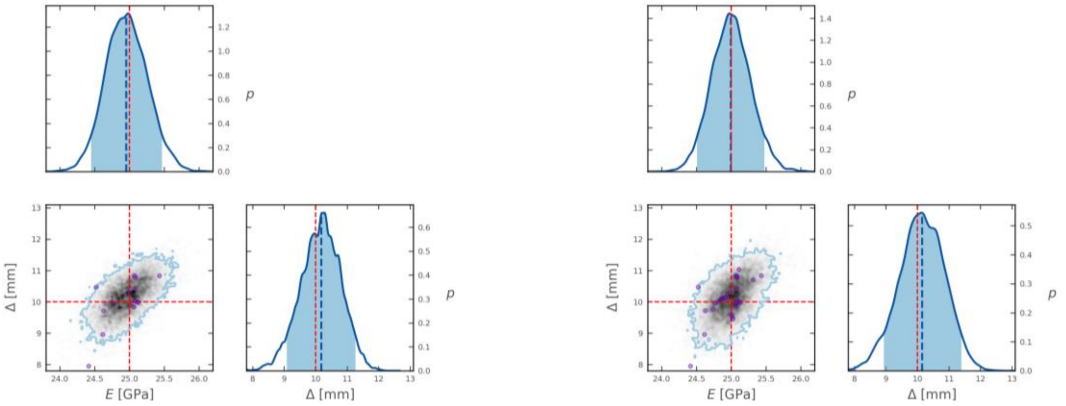
First, a training phase starts by drawing parameter combinations with Latin hypercube sampling and using them as input for evaluating the FE model, after which the log-prior and log-likelihood are obtained as in Section 4.2.1. Then, the GP surrogate is fit using a squared exponential kernel, and its parameters are optimised by maximizing its marginal log-likelihood with equation (30). This last step is repeated a few times to avoid hyperparameters corresponding to local maxima.

For the adaptive phase, three main features need to be defined: the domain of the parameter space, the number of new points to be drawn per step and the convergence threshold ϵ, which represents the change in the mean of the marginal posterior distribution of every parameter relative to the previous marginal posterior distribution standard deviation (see Section 3.5.3). It has to be noted that if convergence is checked, a parameter estimation needs to be performed, and in this case this is done with *emcee* in the same way as in Section 4.2.1, but using the GP surrogate to predict the values associated with the samples. The settings of the BAPE algorithm are presented in Table 5.

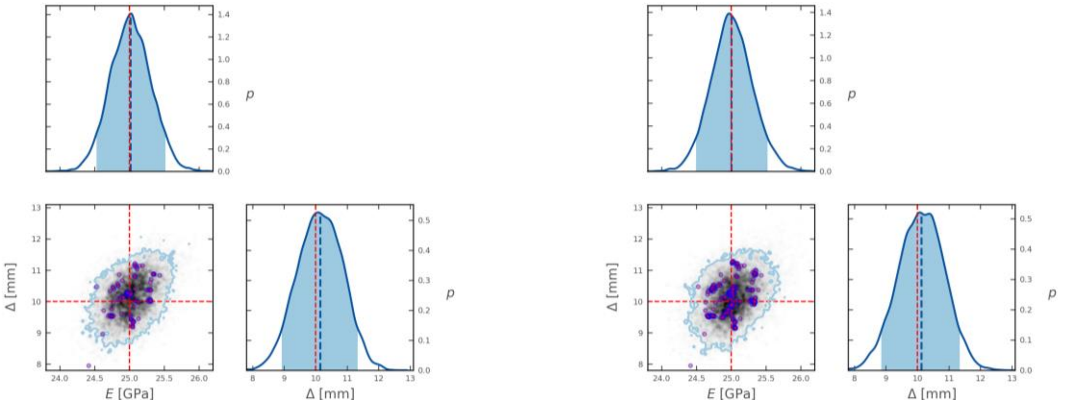
Table 5: Settings of the BAPE algorithm – 2D case.

Initial samples	6
GP Kernel	Anisotropic squared exponential
Number of GP hyperparameter optimisations per step	3
Parameter domain	$E \rightarrow [5, 55]\ GPa$ $\Delta \rightarrow [0, 30]\ mm$
Number of new points drawn per step	4
Convergence threshold $\boldsymbol{\epsilon}$	Multiple values (see Figure 12)
Number of repetitions to reach convergence	3
Number of MCMC walkers	20
Number of MCMC steps	1500

Figure 12 presents the marginal distributions estimated using the BAPE method. It can be seen that the four levels of tolerance provide excellent results since the distributions are shifted towards the ground truth values and the degree of uncertainty is on the same level as the one presented in Figure 11. Also, it is clear that the algorithm is effective on exploiting the high probability density zones (see dots in Figure 12).



(a) $\epsilon = 1.00$, added samples = 16 (b) $\epsilon = 0.1$, added samples = 28



(c) $\epsilon = 0.05$, added samples = 132 (d) $\epsilon = 0.01$, added samples = 400 (*)

Figure 12 (continuation): Posterior distributions obtained with a GP surrogate for different ϵ values – 2D case. The dots represents the samples added by the BAPE algorithm. (*) indicates an analysis without convergence.

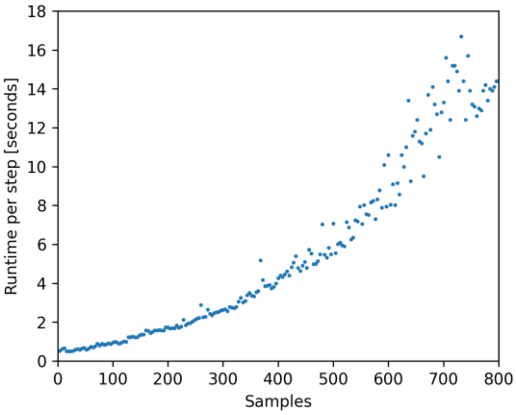


Figure 13: BAPE running time per step (4 added samples each) vs number of samples – 2D case.

It is well-known that one of the main limitations of standard GP regression is its poor scalability when increasing the number of samples, therefore, it is of interest to study the wall-clock time of drawing new samples and fitting the GP as a function of them. Figure 13 shows this for up to 800 added samples. For this plot, the convergence check was not included to avoid distorting the runtimes with the parameter estimations at each step, and also, only one GP hyperparameter optimisation has been performed. The runtime of the FE evaluations is not included either. It should be noted that the runtimes for higher steps appear to be more noisy. A possible explanation for this is that the inherent randomness of hyperparameters optimisation algorithms also scales with the number of samples.

4.3 Description of the problem – 5-dimensional problem

Figure 14 shows the setup of the 5D case. Five parameters are represented as random variables: the Young's modulus of the exterior beams (E_{ext}), middle beam (E_{int}) and columns (E_{col}), as well as the vertical settlement of the left (Δ_1) and right (Δ_2) column base.

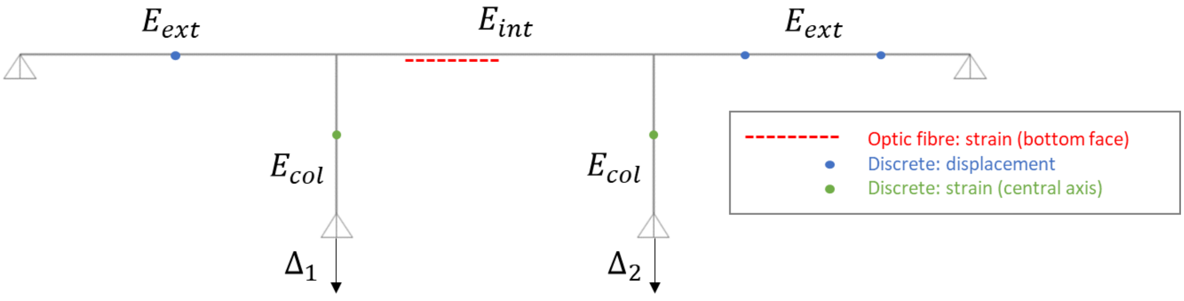


Figure 14: Probabilistic parameters and set of sensors - 5D case.

The same set of sensors from Section 4.1 is used, with the addition of two displacement sensors placed on the right beam at L/3 and 2L/3, and two strain sensors placed on the centre-axis at the middle of both columns.

The FE model prediction is denoted as:

$$\boldsymbol{x}_{\mathrm{FEM}}(E_{\mathrm{ext}}, E_{\mathrm{int}}, E_{\mathrm{col}}, \Delta_1, \Delta_2) = ((\delta 1_1); (\varepsilon 2_1,\ \varepsilon 2_2, \dots, \varepsilon 2_{26}); (\delta 3_1,\ \delta 3_2);\ (\varepsilon 4_1); (\varepsilon 5_1)) \qquad (36)$$

The synthetic measurements are generated in the same way as in Section 4.1 with the parameter values indicated in Table 6.

Table 6: Parameters for synthetic measurements generation – 5D case.

Parameter	Value
E_{ext}	35 GPa
E_{int}	25 GPa
E_{col}	27 GPa
Δ_1	18 mm
Δ_2	12 mm
σ_{δ}	0.2 mm
σ_{ε}	0.5 mm
σ_{model}	0.02
l_{corr}	1000 mm

Table 7 and Figure 15 show the prior distribution.

Table 7: Prior distribution – 5D case. The random variables are assumed to be mutually independent.

Parameter	Prior distribution
E_{ext}	$\mathcal{N}(30, 10)\ GPa$
E_{int}	$\mathcal{N}(30, 10)\ GPa$
E_{col}	$\mathcal{N}(30, 10)\ GPa$
Δ_1	$\mathcal{N}(15, 5)\ mm$
Δ_2	$\mathcal{N}(15, 5)\ mm$

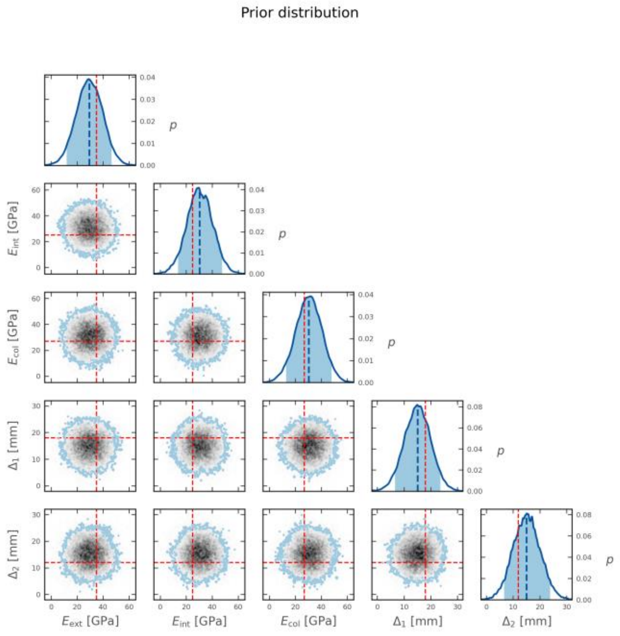


Figure 15: Prior distribution – 5D case. The red and blue dashed lines represent the ground truth values of the parameters and the estimated medians, respectively. Diagonal: 1D marginal densities. Off diagonal: 2D contour plot of joint probability density.

4.4 Results – 5-dimensional problem

In this section, the ground truth parameters are recovered with Bayesian parameter estimation using both a direct MCMC approach and a Gaussian process surrogate approach. The higher dimensionality of this problem allows to check the scalability of the methods proposed.

4.4.1 Direct MCMC estimation

Similarly to the 2-dimensional problem, the parameter estimation of this section is performed with the help of the Python package *emcee* with 50 walkers and 1600 steps.

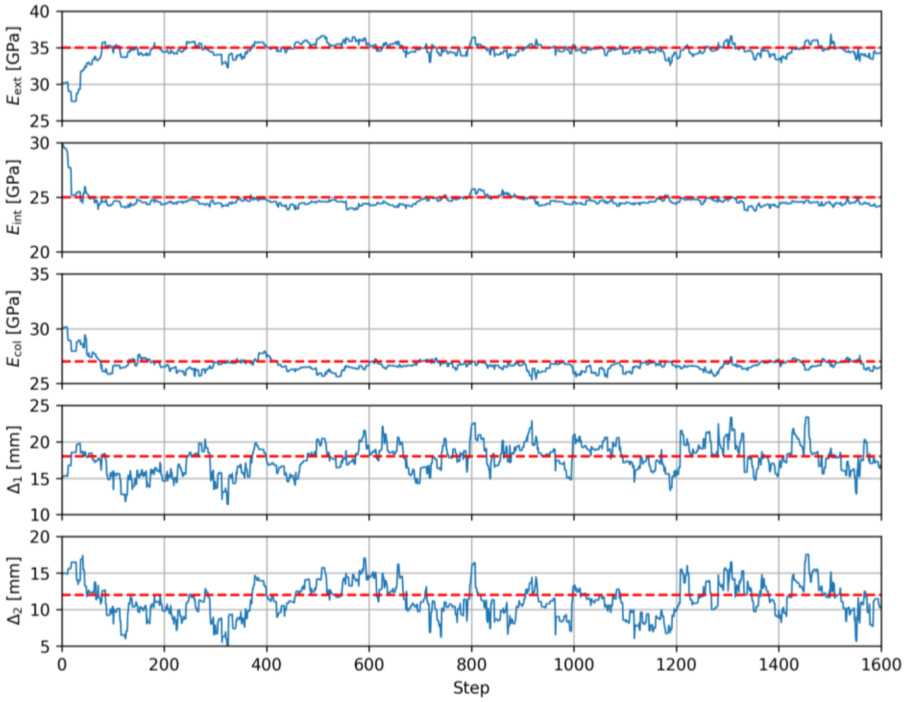


Figure 16: Trace plots for a single walker on 1600 steps – 5D case. The red dashed lines and blue continuous lines represents the ground truth values of the parameters and the samples on every step, respectively.

Figure 16 shows the trace plots of the five physical parameters for a single walker. It can be seen that after approximately 50 steps the walkers start sampling around the ground truth values and independently from the initial positions.

Figure 17 shows the prior and posterior marginal distributions as corner plots, discarding the first 100 steps, while Table 8 presents a summary of the posterior distribution. For the five parameters there is a shift towards the mean values and also a reduction of the standard deviation with respect to the prior, hence the system identification is successful.

Additionally, a strong positive correlation is present between the parameter E_{ext}, Δ_1 and Δ_2. Also, the correlations of E_{int} and the parameters Δ_1 and Δ_2 are moderate and low, respectively. Finally, the parameter E_{col} is virtually independent of the other parameters.

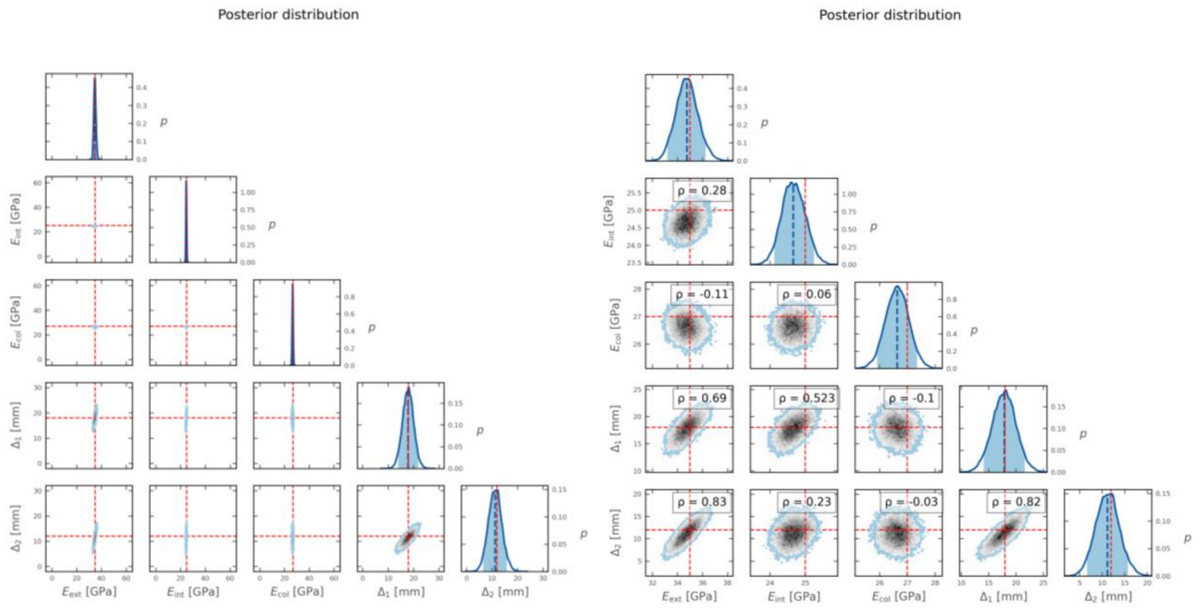


Figure 17: Posterior distribution – 5D case. The left plot shows the same axis limits as the prior while the right one presents a zoomed-in view. The red and blue dashed lines represent the ground truth values of the parameters and the estimated medians, respectively. Diagonal: 1D marginal densities. Off diagonal: 2D contour plot of joint probability density.

Table 8: Summary of the posterior distribution – 5D case

Parameter	Truth	Mean	Median	Std. dev.	25th per.	75th per.	95th per.
E_{ext} (GPa)	35.00	34.78	34.76	0.89	34.18	35.35	36.28
E_{int} (GPa)	25.00	24.67	24.67	0.33	24.45	24.90	25.22
E_{col} (GPa)	27.00	26.64	26.64	0.42	26.35	26.92	27.33
Δ_1 (mm)	18.00	17.86	17.88	2.18	16.39	19.32	21.44
Δ_2 (mm)	12.00	11.26	11.25	2.62	9.47	12.97	15.63

4.4.2 BAPE surrogate model

In the same way as for the 2-dimensional problem, the problem is solved using the BAPE algorithm, whose settings are presented in Table 9.

Table 9: Settings of the BAPE algorithm – 5D case.

Initial samples	15
GP Kernel	Anisotropic squared exponential
Number of GP hyperparameter optimisations per step	3
Parameter domain	$E_{\mathrm{ext}}, E_{\mathrm{int}}, E_{\mathrm{col}} \rightarrow [5, 55]\ GPa$ $\Delta_1, \Delta_2 \rightarrow [0, 30]\ mm$
Number of new points drawn per step	10
Convergence threshold $\boldsymbol{\epsilon}$	Multiple values (see Figure 12)
Number of repetitions to reach convergence	3
Number of MCMC walkers	50
Number of MCMC steps	1500

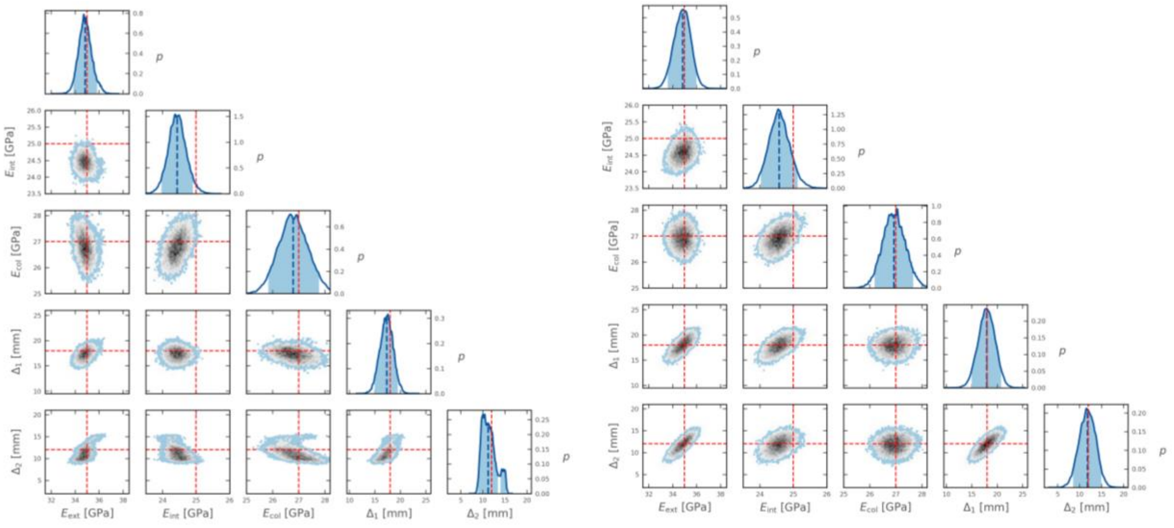
Figure 18 presents the marginal distributions for the 5D case estimated via the BAPE algorithm for three convergence thresholds, while Table 10 and Table 11 are summaries of the mean values and 95% credible intervals obtained with BAPE vs the direct MCMC sampling. From the figures, it is clear the BAPE algorithm is able to capture the shapes, correlations and recover the ground truth values reasonably, even for a threshold of $\epsilon = 1.00$. As expected, the credible intervals and mean values get closer to the direct MCMC results when a more tight threshold is established

Table 10: Summary of the posterior distribution means – 5D case.

Parameter	True val.	MCMC	BAPE, $\epsilon = 1.00$	BAPE, $\epsilon = 0.10$	BAPE, $\epsilon = 0.05$
E_{ext} (GPa)	35.00	34.78	34.30	34.39	34.70
E_{int} (GPa)	25.00	24.67	24.48	24.51	24.63
E_{col} (GPa)	27.00	26.64	27.01	27.03	26.68
Δ_1 (mm)	18.00	17.86	16.03	16.10	17.27
Δ_2 (mm)	12.00	11.26	10.38	10.70	11.00

Table 11: Summary of the posterior distribution 95% credible interval – 5D case

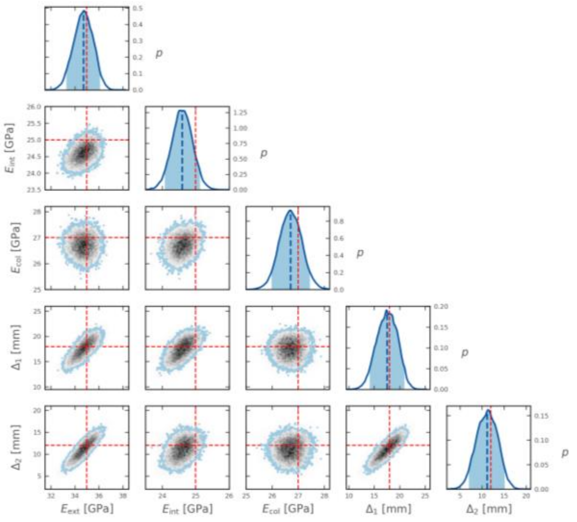
Parameter	True val.	MCMC	BAPE, $\epsilon = 1.00$	BAPE, $\epsilon = 0.10$	BAPE, $\epsilon = 0.05$
E_{ext} (GPa)	35.00	[33.30, 36.28]	[33.17, 35.43]	[33.27, 35.51]	[33.38, 36.02]
E_{int} (GPa)	25.00	[24.12, 25.22]	[23.95, 25.01]	[23.98, 25.04]	[24.08, 25.18]
E_{col} (GPa)	27.00	[25.94, 27.34]	[26.43, 27.59]	[26.43, 27.63]	[25.96, 27.40]
Δ_1 (mm)	18.00	[14.23, 21.49]	[13.58, 18.48]	[13.68, 18.52]	[14.27, 20.27]
Δ_2 (mm)	12.00	[6.89, 15.63]	[7.03, 13.73]	[7.40, 14.00]	[7.43, 14.57]



(a) $\epsilon = 1.00$, added samples = 190

(b) $\epsilon = 0.10$, added samples = 210

Figure 18: Posterior distributions obtained with a GP surrogate for different ϵ values – 5D case. The red dots represents the samples added by the BAPE algorithm. (*) indicates an analysis without convergence.



(c) $\epsilon = 0.05$, added samples = 500 (*)

Figure 18 (continuation): Posterior distributions obtained with a GP surrogate for different ϵ values – 5D case. The red dots represents the samples added by the BAPE algorithm. (*) indicates an analysis without convergence.

Finally, the BAPE runtime per step vs number of samples is presented in Figure 19, in the same way as in Section 4.2.2. In this case, the runtimes for higher steps are noisier as well.

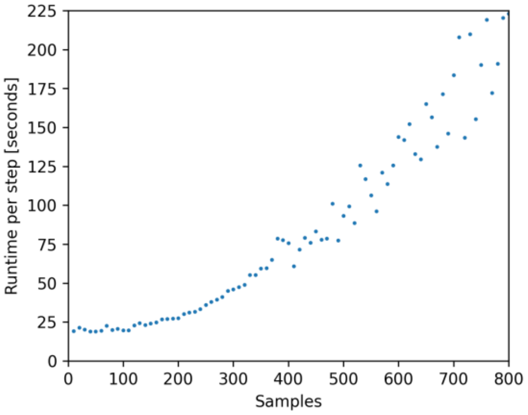


Figure 19: BAPE running time per step (10 added samples each) vs number of samples – 5D case.

4.5 Discussion

4.5.1 Computation time

The BAPE solution presented in Sections 4.2.2 and 4.4.2 proved to be much more efficient than the direct MCMC method in the number of forward model evaluations required; however, it has additional computational costs since it requires the GP surrogate to be fitted at every step, as well as the selection of new training points. In order to compare the two methodologies in a more extended way, the total runtime for the analyses presented in the previous sections are calculated by considering different runtimes per forward model evaluation. The considered runtimes range from 10^{-3} to 10^{4} seconds. The results of this comparison are shown below.

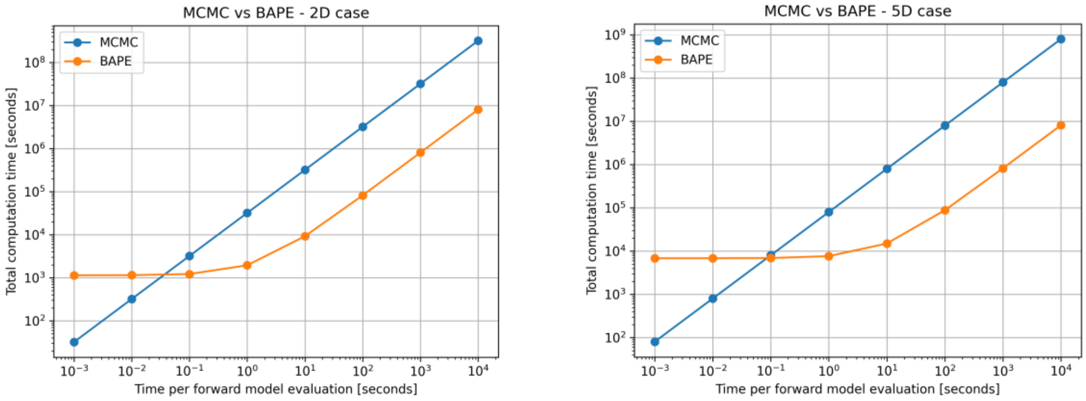


Figure 20: Comparison of MCMC and BAPE total runtimes as a function of the time per forward model evaluation for the 2D case (left) and 5D case (right). For this case, the real runtime for a single finite element evaluation is approximately 0.002 seconds.

As expected, the direct MCMC method has a lower computational time only for very fast forward models. Indeed, the BAPE algorithm becomes more efficient at 0.02 and 0.10 seconds for the 2D and 5D cases, respectively, which is way below the runtimes of realistic civil engineering structural analyses performed in commercial software. In addition to that, it can be observed that at around 10 seconds for the left graph and 100 seconds for the right graph, the BAPE runtimes start being governed almost exclusively by the forward model runtime; however, higher dimensional problems may require a larger number of samples which could make the GP training the bottleneck if the FE model runtime is in the realm of a few minutes. In such case, a fast approximate GP formulation may have to be used.

Finally, the default option of *approxposterior* is to verify convergence by performing a parameter estimation at every step with an MCMC approach using the GP surrogate. This considerably slows every step so it was removed from the runtimes shown in this sub-section. Strategies such as checking convergence only every '*n*' number of steps or estimating the means and standard deviations by approximate Bayesian methods rather than MCMC, could be used to reduce the burden of convergence checks on the total computation time.

4.5.2 GP performance

A typical anisotropic squared exponential kernel was used for the GP surrogates of this chapter since it requires relatively few hyperparameters to tune. Another advantage of this kernel is that it is *universal* (Micchelli, et al., 2006), which means that it is capable of uniformly approximating any continuous function with enough data. Duvenaud (2014) states that sometimes this kernel flexibility comes at the cost of slowness, and that accounting for a more complex kernel structure generally means less data needed, i.e., the *blessing of abstraction* (Goodman, et al., 2011) that counters the *curse of dimensionality*. This implies that there is potential of further reduction of the number of training points with an alternative kernel structure.

Nevertheless, for the purpose of this thesis, no further investigation is done with respect to kernel structures, especially considering that the one chosen was able to appropriately approximate the log joint probability with a reasonably low number of training points for the

2D and 5D cases. Additionally, the results shown in Sections 4.2.2 and 4.4.2 did not vary considerably when the hyperparameter optimisation was performed once instead of three times every step.

It should be noted that the accuracy of the GP surrogate is less relevant during intermediate steps since then it is used only to draw new training points. According to Wang & Li (2017) it is not essential for them to be the optimal solution as long as good training points are found. In contrast, for the last step the accuracy of the GP becomes very important since here it is used for the MCMC-based parameter estimation, therefore, in this case a higher number of hyperparameter optimisation restarts may be recommendable.

Finally, there might be an opportunity of improvement in the number of training points needed if the algorithm that maximizes the utility function can be modified in a way that penalizes the location of points too close to each other, thus, avoid some of the clusters that can be seen in Figure 12.

4.5.3 Calculation of posterior predictives and evidence

The focus of this chapter is on obtaining reliable and cheap-to-compute posterior distributions. However, often it is of interest to also predict the probability distribution of future observations given the existing set of observations. These posterior predictive distributions can be calculated with the following equation:

$$p(\tilde{\boldsymbol{y}}|\boldsymbol{y}) = \int p(\tilde{\boldsymbol{y}}|\boldsymbol{\theta}).\,p(\boldsymbol{\theta}|\boldsymbol{y}).\,d\boldsymbol{\theta} \tag{37}$$

where $\tilde{\boldsymbol{y}}$ is the vector of future observations and y is the vector of the already existing observations. The second term of the integral is the posterior distribution which in this case has been obtained by an MCMC method, using a GP rather than directly calling the forward model. In theory, new samples may not be needed to estimate posterior predictives since the ones drawn for MCMC can be used. The problem is that the future observations likelihood term $p(\tilde{\boldsymbol{y}}|\boldsymbol{\theta})$ requires the evaluation of the forward model at every sample. Nevertheless, if the responses at the sensor locations and the prediction locations are stored every time the FE model is evaluated under the BAPE algorithm, then a GP surrogate of a second likelihood function can be constructed for the posterior predictive using the prediction locations. A different convergence criteria may be necessary for this second surrogate in order to check its accuracy and some extra new training points may be needed.

On the other hand, an accurate value of the evidence under the BAPE approach may not be easy to obtain even if the parameter space is sampled in an evidence-driven way like Nested sampling (Skilling, 2006), since the GP has been trained only in the areas that highly contribute to the posterior distribution accuracy. In this case, a procedure that actively learns the evidence by means of Bayesian Quadrature (Osborne, et al., 2012) may be preferable.

5 Case study 2: Bridge 705 - Description

In this chapter, the real world case study Bridge 705 is presented. The geometry of the bridge is described in Section 5.1; while Section 5.2 introduces the measurement campaigns conducted by several parties, both with discrete sensors and optic fibre sensors. The finite element model of the structure developed in the structural engineering software DIANA FEA (2022) is presented in Section 5.3. Moreover, the parameter estimation sub-cases and the specific goal behind each of them are discussed in Section 5.4. Finally, the settings for parallelization are presented in Section 5.5. It should be noted that some Bayesian parameter estimations have previously been performed for this bridge by Rózsás, et al. (2022).

5.1 Structure

Bridge 705 is a reinforced concrete bridge located in the western part of Amsterdam which spans over the Johan Huizingalaan street and was inaugurated in 1962. Its central part carries tram lines 1, 17 and 27, while the sides are destined to road lanes. Figure 21 shows recent pictures of the bridge

Figure 21: Overview of Bridge 705 (Amsterdam op de kaart, 2022).

The bridge consists of a 7-span continuous reinforced concrete deck supported by 24 intermediate V-shaped pillars aligned in 6 rows and connected to the foundations with hinges. The deck is monolithically connected to the pillars, and its thickness varies in the transverse direction from 45cm to 66cm, while it remains constant longitudinally. Figure 22 shows a cross-section view of the bridge. Figure 23 shows top and side views of the bridge.

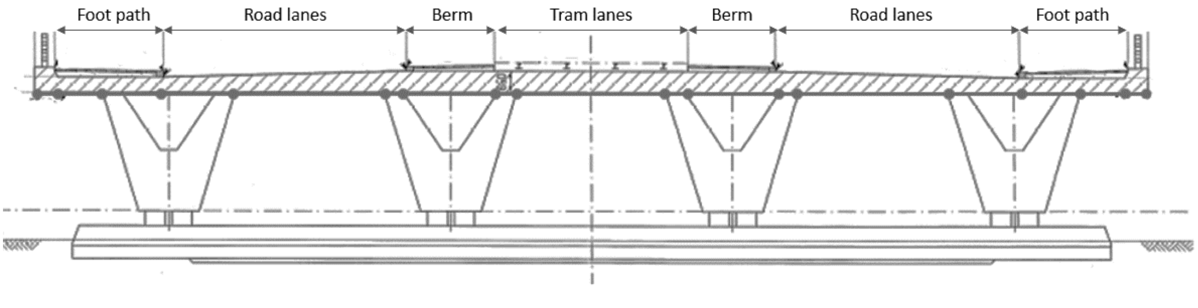


Figure 22: Cross-section view of Bridge 705 (Rózsás, et al., 2022).

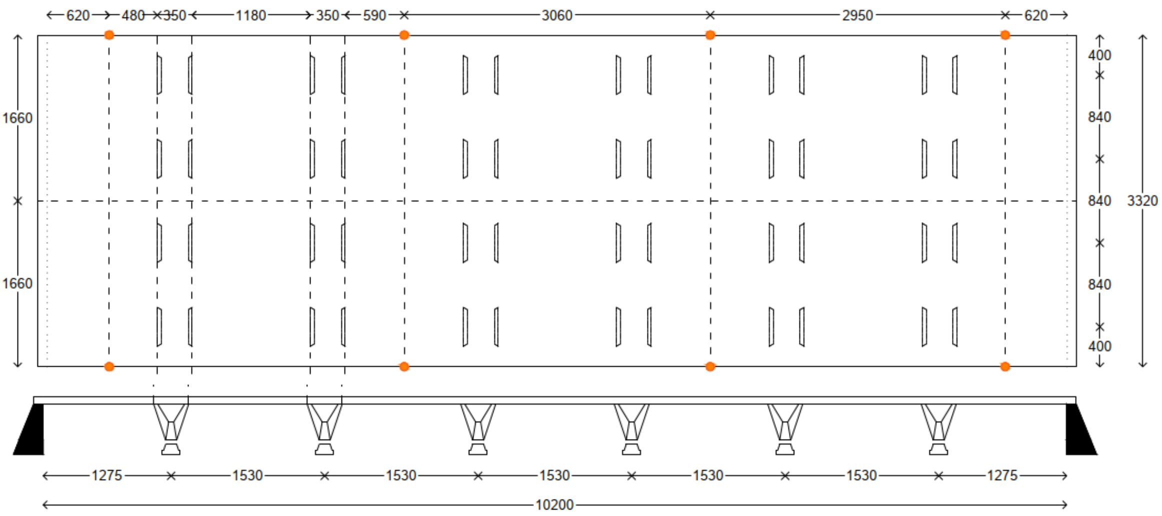


Figure 23: Top and side longitudinal view of Bridge 705. Dimensions are in cm (Rózsás, et al., 2022).

5.2 Measurement campaign

In October 2018, TNO and several other organizations conducted a measurement campaign in Bridge 705. For a couple of nights, traffic was closed and two sand trucks were positioned in different configurations, which are shown in Figure 24.

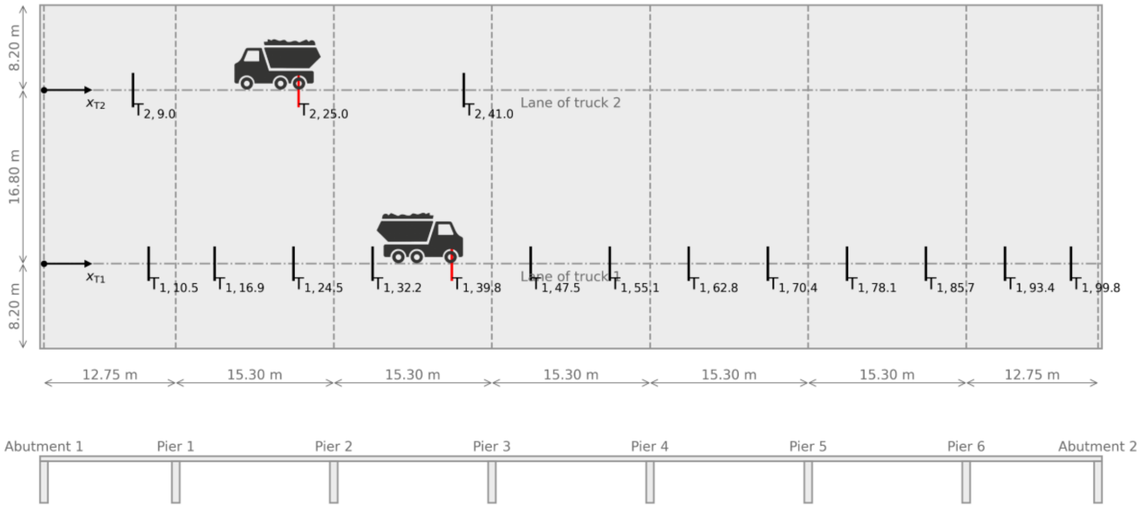


Figure 24: Axle positions of Truck 1 and Truck 2 used for measurement campaign. Top view (top) and side view (bottom) of the bridge (Rózsás, et al., 2022).

Truck 1 had three axles and weighted around 35 ton, while Truck 2 had four axles and weighted 45 ton. Before the start of campaign, the exact axle load and axle distances were measured.

Four static test series are considered, accounting for a total of 52 load positions:

1. Series 'ZB': Truck 1 in each of its 13 positions without Truck 2
2. Series 'MB1': Truck 1 in each of its 13 positions with Truck 2 in position 1

3. Series ‘MB2’: Truck 1 in each of its 13 positions with Truck 2 in position 2
4. Series ‘MB3’: Truck 1 in each of its 13 positions with Truck 2 in position 3

Figure 25 and Figure 26 show the layout of discrete translation and strain sensors used in the campaign, with the different labels corresponding to the organization that installed them. In all cases, the sensors are installed under the bottom face of the deck.

For each sensor and load case, a single static measurement value is obtained by averaging the results during the one minute that the truck is in position. Additionally, a spatial averaging was applied for the strain sensors since the strains are not obtained on a single point, but along the sensor extension (see Table 12).

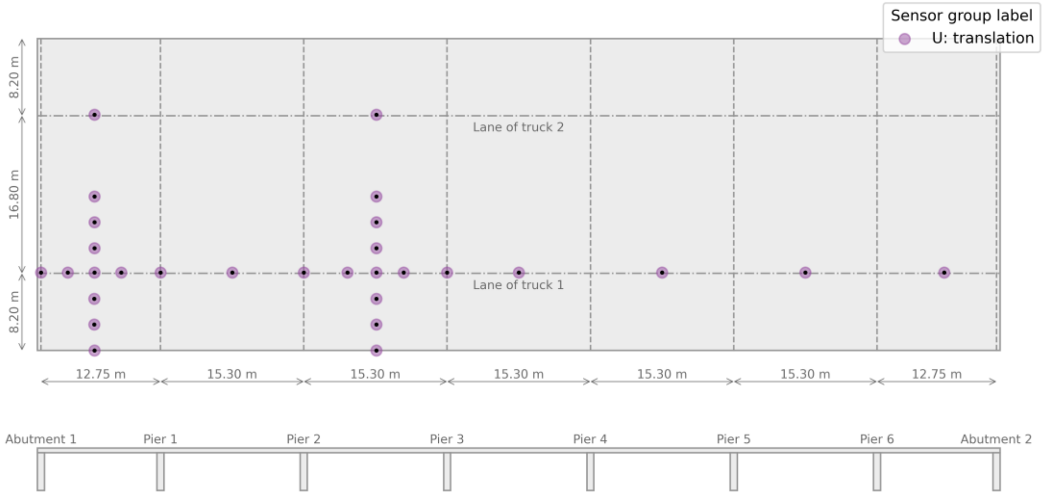


Figure 25: Location of discrete translation sensors used for measurement campaign. Top view (top) and side view (bottom) of the bridge (Rózsás, et al., 2022).

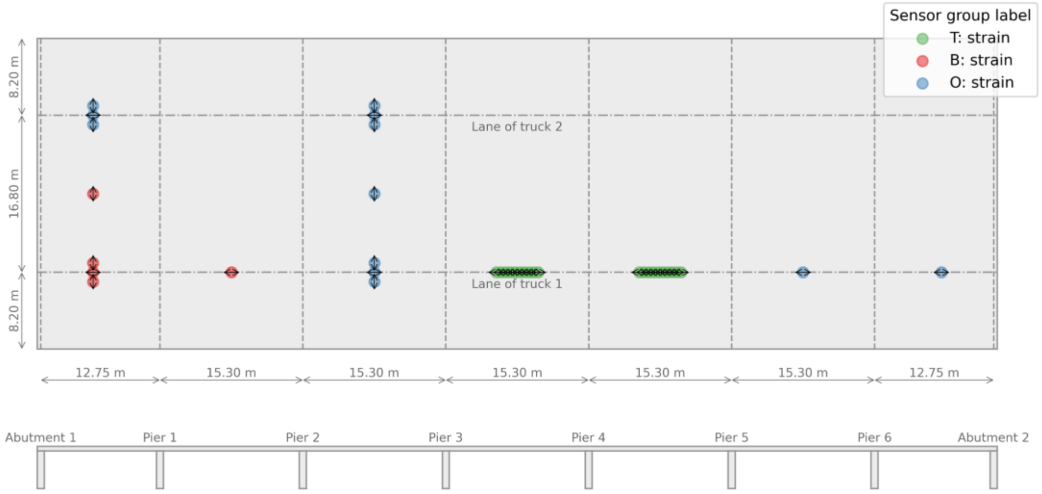


Figure 26: Location of discrete strain sensors used for measurement campaign. Top view (top) and side view (bottom) of the bridge. The arrows indicate the direction of the measured strain (Rózsás, et al., 2022).

Table 12: Discrete sensor groups outline.

Sensor group	Measured quantity	Number of sensors	Sensor length
U	Translation (vertical)	29	N. A.
B	Strain	20	2.00 m
O	Strain	12	2.00 m
T	Strain	5	0.50 m

In addition to the sensor groups mentioned before, a novel high resolution optic fibre sensor (Wosniok, et al., 2019) was also installed, giving strain measurements along its path with a relatively high spatial resolution (at approximately every 10 cm). The layout of the fibre and some of its properties are presented in Figure 27 and Table 13.

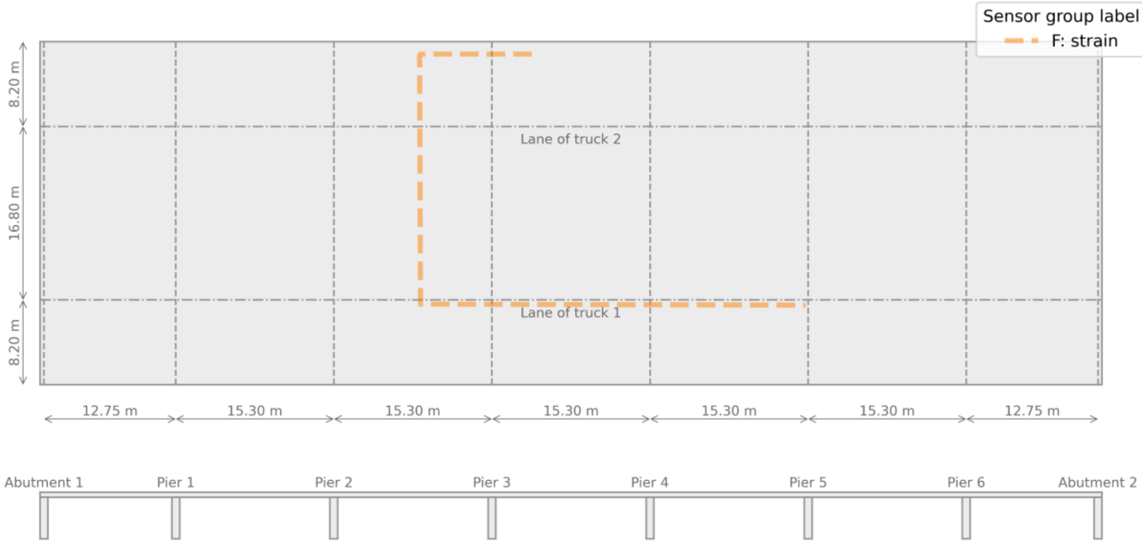


Figure 27: Location of optic fibre strain sensor used for measurement campaign. Top view (top) and side view (bottom) of the bridge.

Table 13: Optic fibre outline.

Fibre leg	Strain orientation	Leg length	Discretization points
Top	X	10.85 m	106
Transversal	Y	24.30 m	236
Bottom	X	34.50 m	336

More information about the measurement data processing and the technology behind every sensor group can be found in Rózsás, et al. (2022).

5.3 Finite element model

This thesis uses a 3D finite element model previously developed by Rózsás, et al. (2022) in the software DIANA FEA, to predict Bridge 705 responses. The deck and pillars are modelled with quadrilateral 8-noded shell elements with different thicknesses and a mesh size between 20cm and 25cm. The elastic modulus and Poisson ratio are the same for all concrete in the bridge; however, the contributions of the asphalt layer and tram tracks on the stiffness of the deck are taken into account as equivalent elastic modulus with fixed values for the pavement and rail material properties.

The hinges at the base of the pillars are considered as nodal rotational springs in the longitudinal direction, while fully restricting translations in the vertical and transverse directions. The deck supports at the abutments are modelled as rollers with free longitudinal translations.

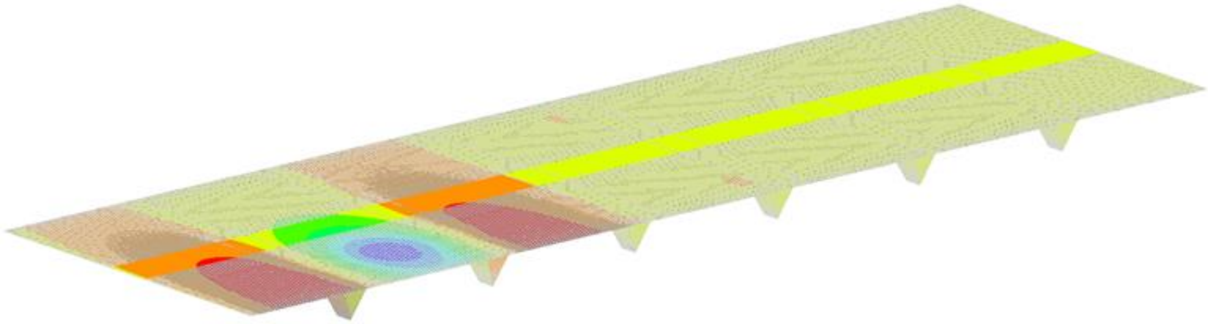


Figure 28: 3D view of finite element model. The contours show the response under a concentrated load in span 2 (Rózsás, et al., 2022).

Given the low load intensity of the test trucks, all analyses are linear elastic. The self-weight of the structure is not considered since the sensors were installed after all bridge dead loads have been applied. The truck loads are modelled as uniformly distributed loads on a rectangular area of 90 x 96cm which represents the 45° projection of the tires footprints at the centre of the deck thickness.

The movement of trucks 1 and 2 is discretized every meter and the truck locations in between are linearly interpolated. The translation predictions correspond to the node closest to the sensor location, while the strains are obtained by averaging the centroidal strains for the elements crossed by the sensor path.

5.4 Parameter estimation sub-cases

The followed approach for the Bayesian parameter estimations of Bridge 705 is the one described in Section 3.5.3. In contrast with the previous case study, an *ad-hoc* Python code developed by the author is used instead of *approxposterior*, in order to have more flexibility to introduce the cloud-based parallelization scheme described in Section 5.5.

With the measurement data described in 5.2, four main sub-cases are investigated, as shown in Table 14. For all of them, two physical parameters are considered to be identified: the concrete elasticity modulus E_c and the springs rotational stiffness K_{rot}. Table 15 and Figure 29 show the selected prior distribution.

Table 14: Description of parameter estimation sub-cases.

Case	Parameters	Sensor subsets	Test Series
Sub-case 1	E_c, K_{rot}	‘B’, ‘O’, ‘T’ and ‘U’	‘ZB’
Sub-case 2	E_c, K_{rot}	‘O’ (partially) and ‘T’	‘ZB’
Sub-case 3	E_c, K_{rot}	Transversal and bottom optic fibre legs	‘ZB’
Sub-case 4	E_c, K_{rot}, σ_{meas}, cov_{model}, L_{corr}	Transversal and bottom optic fibre legs	‘ZB’

Table 15: Prior distribution of physical parameters.

Parameter	Distribution	Lower bound	Upper bound
E_c (GPa)	Uniform	10	100
K_{rot} (Nmm/rad)	Uniform	1	10^{12}

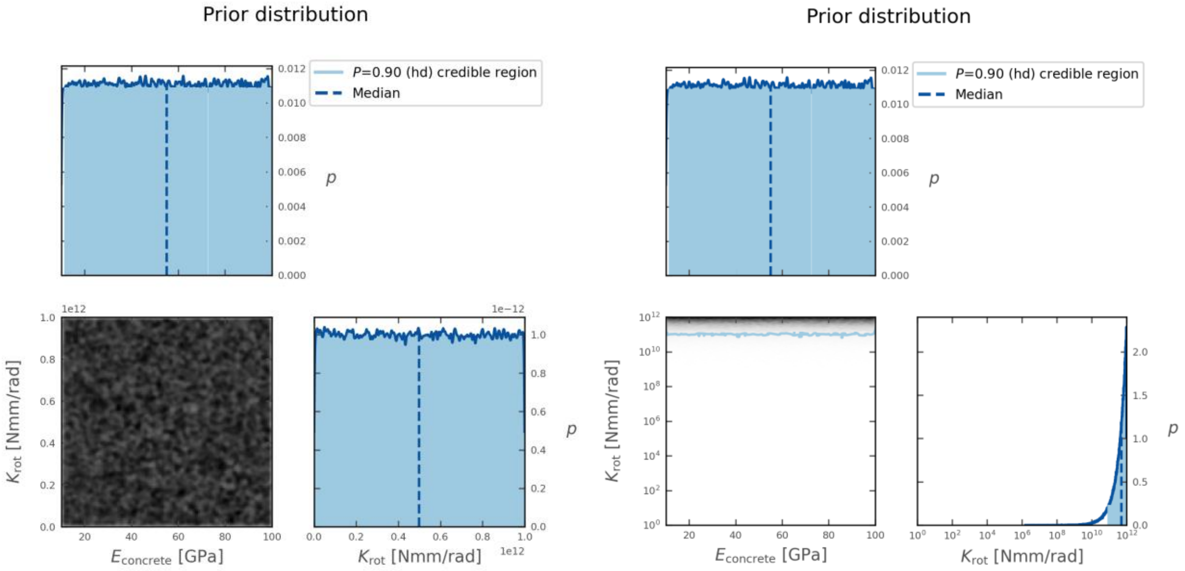


Figure 29: Prior distribution of physical parameters. The plot on the left shows K_{rot} in linear scale while the plot on the right shows it on log-10 scale. Diagonal: 1D marginal densities. Off diagonal: 2D contour plot of joint probability density.

As described in Section 3.2, a correlated model prediction uncertainty (Table 16) and an uncorrelated measurement uncertainty (Table 17) are considered. These values were also used by Rózsás, et al. (2022), who estimated them based on expert judgement.

Table 16: Fixed model prediction uncertainty parameters for sub-cases 1, 2 and 3.

Group	cov_{model}	L_{corr} (m)
Translation	0.05	3.00
Strain	0.10	3.00
Load position	NA	3.00

Table 17: Fixed measurement uncertainty parameters for sub-cases 1, 2 and 3.

Sensor group	σ_{meas}
U, translation	0.05 mm
B, strain	1.00 microstrains
O, strain	1.00 microstrains
T, strain	0.50 microstrains
Optic fibre, strain	1.00 microstrains

All first three sub-cases start with 20 initial samples drawn from the parameter domain with Latin hypercube sampling, as shown in Figure 30.

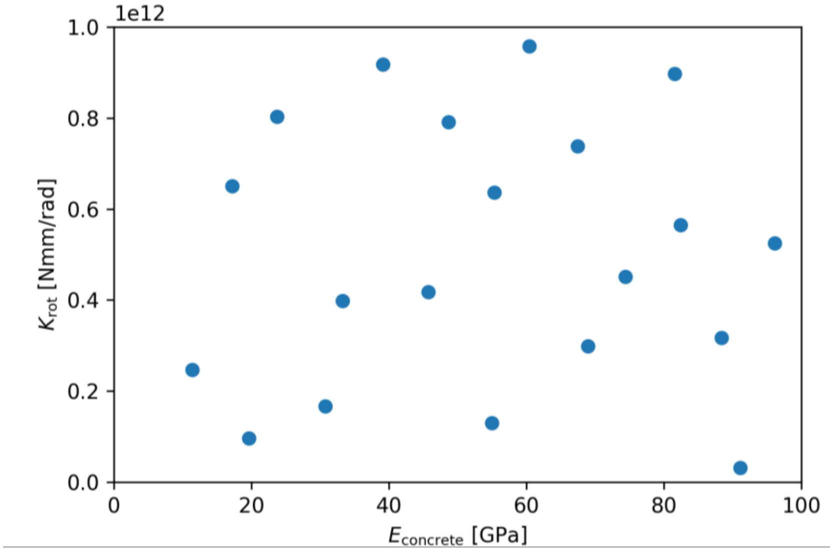


Figure 30: Initial Latin hypercube samples used for sub-cases 1, 2 and 3.

Sub-case 1 makes use of all discrete strain and translation sensors, and serves to validate the thesis methodology against Rózsás, et al. (2022), which has the same probabilistic settings but performs a conventional MCMC sampling, i.e., without using of a log joint probability surrogate. Nevertheless, it should be noted that they did use a GP surrogate to reduce the computation time, but on the finite element model predictions. This structural GP[11] was trained with a grid of 112 points over the parameter domain.

Sub-cases 2 and 3 use subsets of discrete and optic fibre sensors, respectively, with the goal of finding what gain in information is obtained when using the latter in comparison with the former.

On top of the two physical parameters, sub-case 4 also aims to identify the three uncertainty parameters that were fixed for the previous sub-cases, in order to assess their influence in the context of Bayesian parameter estimations.

---

[11] Technically, it is a collection of many individual GPs, one for each sensor and load case. No correlation structure between outputs is included

Gamma distributions are assumed for the priors, following the recommendation from Simoen, et al. (2013). The probability densities for the three parameters are shown in Figure 31 and the full prior is shown in Figure 32. The Gamma distribution parameters are presented in Table 18 and they have been selected such that their means coincide with the fixed values from sub-case 3 while moderately limiting them on the right side. This constraint is introduced to avoid excessively high values (e.g. a very large σ_{meas} that causes all measurements to be identified as noise), although it is possible that this might be too biased.

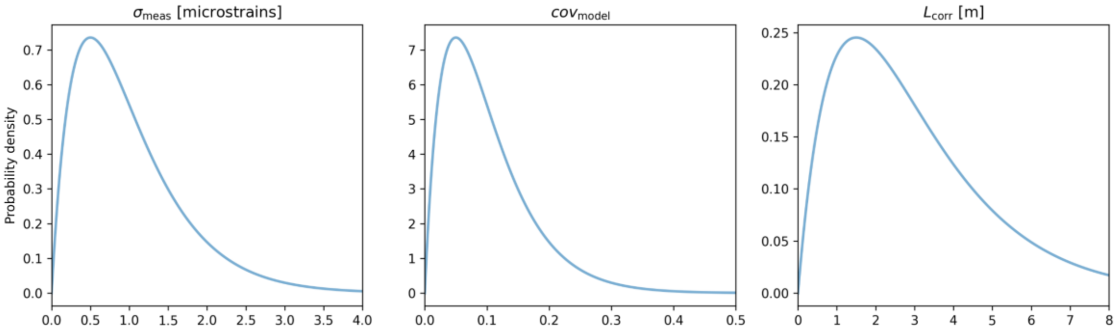


Figure 31: Prior probability density for uncertainty parameters.

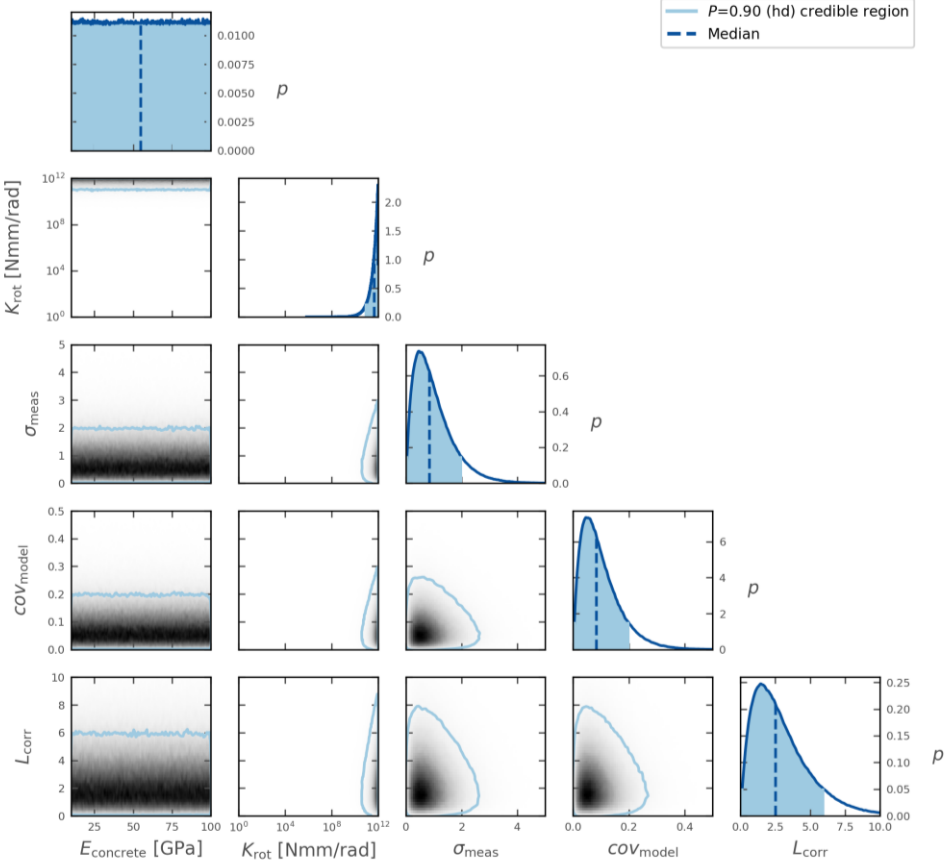


Figure 32: Prior distribution of physical parameters and statistical parameters. K_{rot} is plotted on log-10 scale. Diagonal: 1D marginal densities. Off diagonal: 2D contour plot of joint probability density.

Table 18: Prior distribution of uncertainty parameters for sub-case 4.

Parameter	Distribution	k (shape)	θ (scale)
σ_{meas}	Gamma	2.00	0.50
cov_{model}	Gamma	2.00	0.05
L_{corr}	Gamma	2.00	1.50

In addition to the 20 Latin-hypercube points from Figure 30, four physical parameter combinations corresponding to the corners of the domain are added, since some tests showed that the GP mean predictions were severely overestimated in these regions, making the algorithm draw new points there even though they did not contribute much to the posterior distribution accuracy.

With the finite element analysis results already available, it is not computationally demanding to calculate the log joint probabilities for different statistical parameter combinations. In this case, two grids were used, giving 8 and 27 points, respectively, as shown in Figure 33. The 24 physical parameter points are combined with the 35 statistical points, giving a total of 840 samples for Sub-case 4. The data is then split and 75% is used to start the posterior active learning procedure, while the remaining is employed to test the accuracy of the GP.

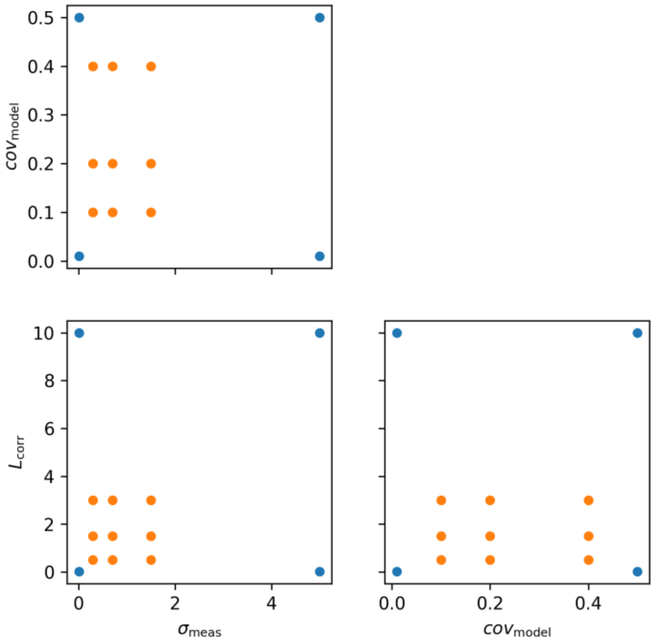


Figure 33: 2D visualization of statistical parameters grid points. The blue and orange dots correspond to Grid 1 (8 points) and Grid 2 (27 points), respectively.

5.5 Parallelization settings

The cloud-based parallelization tool described in Section 3.6 is used in this case study to evaluate multiple finite element models simultaneously, thus reducing the Bayesian parameter estimation wall-clock time. Table 19 shows the specifications of the Azure virtual computer selected to run the DIANA FEA.

Table 19: Specifications of used Azure virtual computers

Name	Standard_D16a_v4
Number of vCPUs	16
Memory	64 GiB
Processor	AMD EPYC 7452 32-Core processor
OS disk size	1023 GiB
Resource disk size	400 GiB

The adaptive sampling algorithm selects new points by optimizing a utility function, so theoretically, only one point could be drawn per step and parallelization would not be possible. To circumvent this problem, a distance constrain is introduced for this case study such that if a proposed point is too close to either any of the existing points or the other candidates, the point is rejected. Indeed, if it is found that distances with any point are below the established distance thresholds for every parameter, the utility function returns a negative infinite, forcing the optimization algorithm to explore in other areas. On the other hand, if the distances are fulfilled, the utility function returns the *exponentiated variance* presented in Section 3.5.3.

Table 20 shows the distance thresholds used for this case study. The values were chosen to allow for exploitation of high probability areas while avoiding the dense clustering of points observed in Chapter 4. Ideally, these settings should be tuned to maximize the algorithm efficiency; however, this aspect is beyond the scope of this thesis and is not addressed further.

Table 20: Parallelization distance thresholds for physical and statistical parameters

Parameter	Distance threshold
E_{c} (GPa)	2.0
K_{rot} (Nmm/rad)	5E+10
σ_{meas}	0.5
cov_{model}	0.2
L_{corr} (m)	1.0

6 Case study 2: Bridge 705 - Results

In this section, the results of the sub-cases introduced in 5.4 are presented. For all of them, the converged posterior distributions are displayed. In addition to that, since an active learning procedure is used to approximate the log joint probability density, it is also of interest to investigate the evolution of the posterior evolution per step and the distribution of the new samples. Moreover, the finite element predictions obtained with the posterior median parameters are compared with the a priori model predictions, in order to assess the Bayesian parameter estimation efficacy. Finally, the influence of the measurement and model uncertainty parameters is evaluated for the optic fibre case.

6.1 A priori model

The a priori modulus of elasticity is calculated based on compressive tests applied on concrete cores drilled from the bridge deck, while the stiffness of the rotational springs at the piers base is set to 1.0 Nmm/rad, hence simulating an ideal hinge. The values are shown in Table 21.

Table 21: A-priori parameter values.

Parameter	A-priori value
E_c (GPa)	38.4
K_{rot} (Nmm/rad)	1.0

A comparison of the finite element predictions with the measurements for the four discrete sensor groups is presented in Figure 34. It can be seen that a reasonable agreement is obtained for both translations and strains. Note that in case of perfect agreement all dots are at the diagonal line and the coefficient of determination (R^2 score) is equal to 1.0.

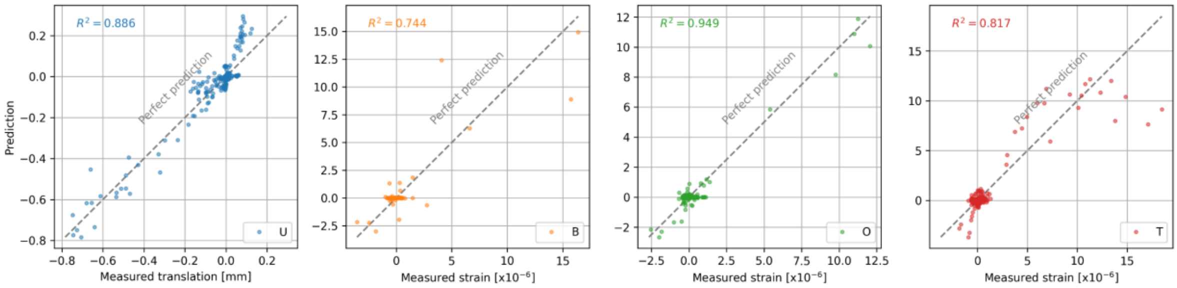


Figure 34: Comparison of a priori model predictions (y-axis) vs measurements (x-axis) for translation ('U') and strain ('B', 'O' and 'T') sensor groups.

In a similar way, the predictions and measurements are compared for the optic fibre sensor in Figure 35. Here the agreement is also good for the a priori model. It should be noted that some corrections have been applied to this set of measurements, as described in Appendix B.

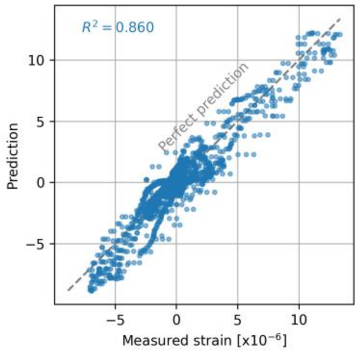


Figure 35: Comparison of a priori model predictions (y-axis) vs measurements (x-axis) for the optic fibre sensor.

Figure 36 provides more insight in the discrepancy between predictions and measurements, showing longitudinal strain diagrams for two fixed truck locations. It can be seen that the shape of the optic fibre measurements is quite similar to the model predictions, although with slight differences in the peaks. On the other hand, the discrete strain measurements show a high variability between consecutive points (with a 0.5m distance) and also with respect to the predictions.

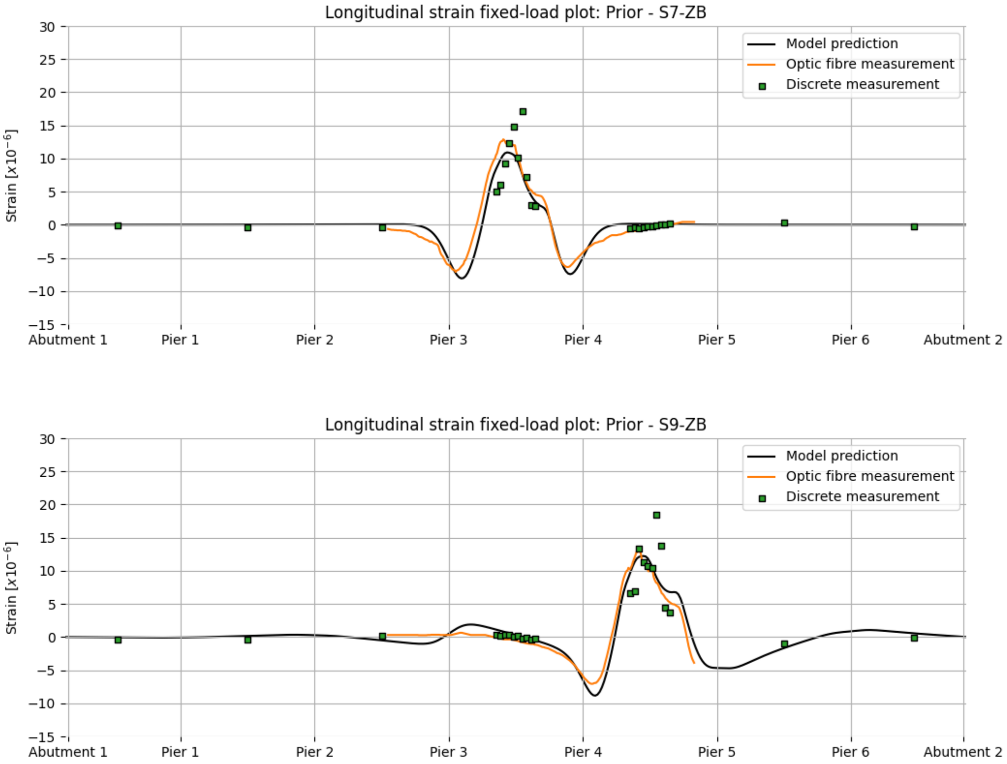


Figure 36: Longitudinal strain diagrams at centreline of Truck 1 lane (deck bottom fibre). The model predictions are obtained with the a priori model parameters. Top: load case S7-ZB. Bottom: load case S9-ZB.

In a similar way, Figure 37 shows vertical translation diagrams for the same truck positions. For both load cases, there is a non-negligible difference between the measured and predicted peaks. Moreover, the predicted deck rotations above the pier lines are underestimated for load case S7-ZB and overestimated for S9-ZB.

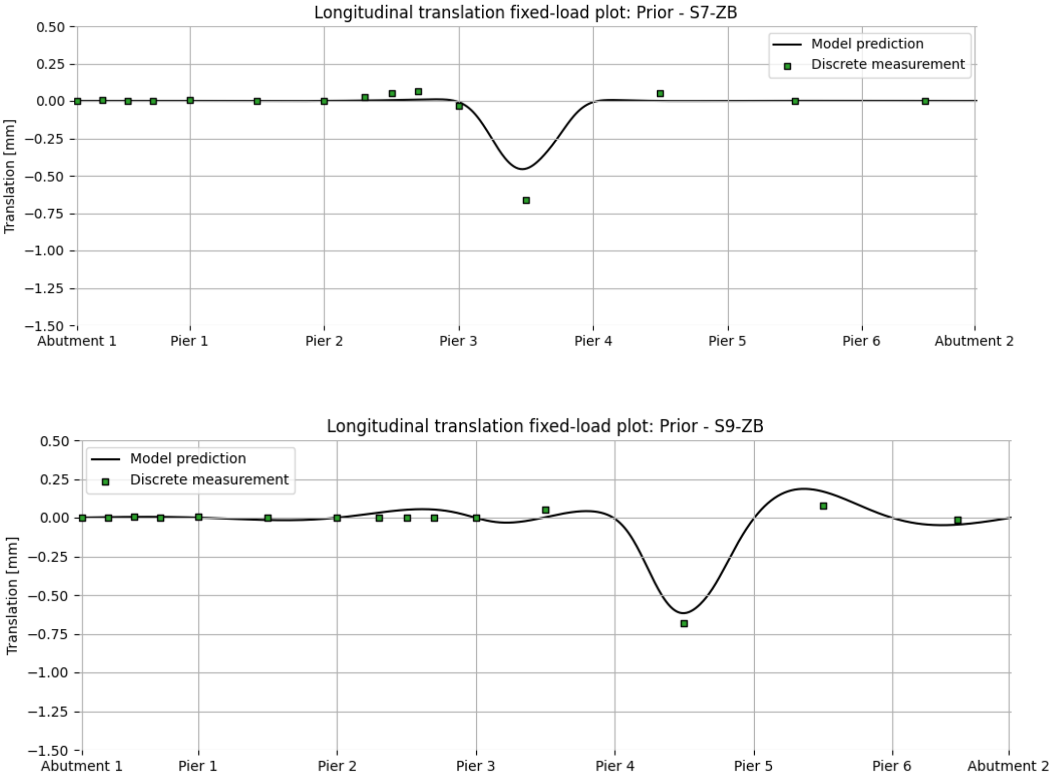


Figure 37: Vertical translation diagrams at centreline of Truck 1 lane. The model predictions are obtained with the a priori model parameters. The model predictions are obtained with the a priori model parameters. Top: load case S7-ZB. Bottom: load case S9-ZB.

6.2 Sub-case 1

Sub-case 1 uses all Series ZB discrete measurements to perform the Bayesian parameter estimation. The active learning is started with 20 initial samples and 4 new samples are drawn per step based on an exponentiated variance utility function. When convergence is achieved, a final MCMC sampling using 20 walkers and 15,000 steps is performed in order to obtain the posterior medians and median absolute deviations. The results are summarized in Table 22. For the concrete elastic modulus E_c a reduction of around 28% is obtained with respect to the a priori value. On the other hand, the change of the rotational stiffness is quite significant, showing an almost clamped behaviour while the a priori model assumed ideal hinges.

Table 22: Posterior values for Sub-case 1 – Log joint probability GP.

Parameter	Median	MAD
E_c (GPa)	27.48	0.65
K_{rot} (Nmm/rad)	8.80 E+11	1.16 E+11

The converged posterior distribution is shown in Figure 38 using a log-10 scale for K_{rot}. It is evident when comparing the left sub-figure with the prior distribution from Figure 29 that a significant concentration of the probability densities on the top left corner has occurred. It is also noticeable from the right sub-figure that the densities for K_{rot} are tilted towards the right boundary of the domain, which is even more noticeable on a logarithmic scale. Even though is not ideal to have such behaviour, the K_{rot} boundary has not been extended since the effect of having higher values is minimal on the model predictions, i.e., the hinges at the bottom of the piers already behave very similar to a clamped support.

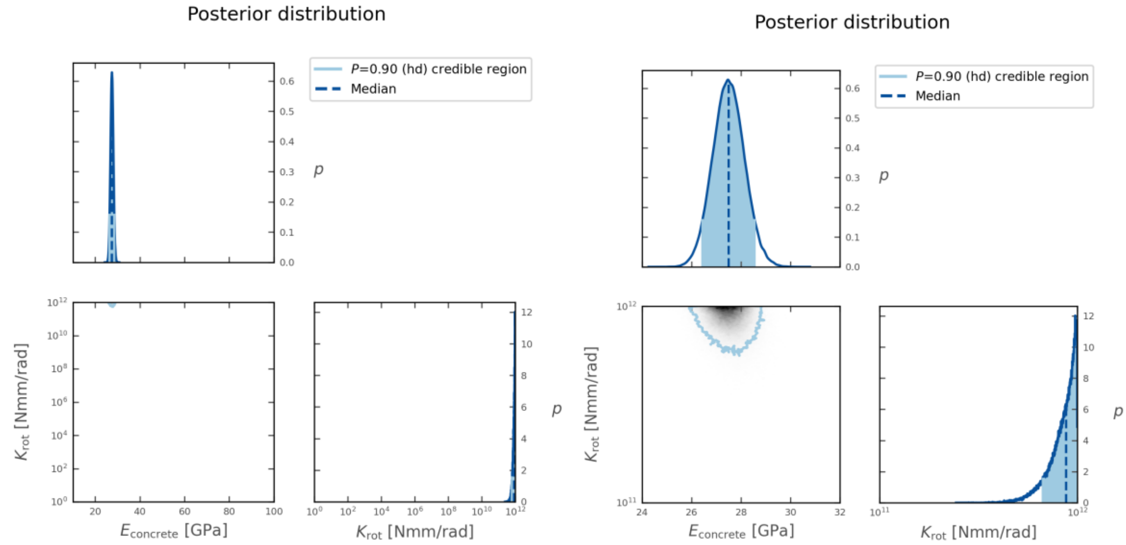


Figure 38: Converged posterior distribution for Sub-case 1 obtained by directly using FE model predictions. The left plot shows the same axis limits as the prior while the right one presents a zoomed-in view. Diagonal: 1D marginal densities. Off diagonal: 2D contour plot of joint probability density.

In order to validate the active learning procedure proposed in this thesis, the Bayesian parameter estimation is also performed with a direct MCMC sampling, meaning that the log likelihood is calculated for every sample according to equation (10), rather than predicted with a log probability GP. This means that strain and translation predictions are needed at the location of all sensors and load cases for every parameter value. Since it is unfeasible to run hundreds of thousands of finite element models[12] a structural multi-output GP surrogate built with 112 training points is used to predict the bridge responses for every sensor at a negligible computing cost. More details about this can be found in Rózsás, et al. (2022), whose Python code was used to obtain the GP structural predictions.

Table 23 shows the posterior medians and median absolute deviations using the structural GP. The posterior values for the concrete elastic modulus E_c show a very good agreement with the ones from Table 22, having a difference of 0.18 GPa for the median and 0.11 GPa for the MAD. The differences are more visible for the rotational stiffness K_{rot}, with 0.86E+11 Nmm/rad for the median and 0.64E+11 Nmm/rad for the MAD. It should be noted that both median predictions are within less than one MAD from the other and that variability in model predictions for high K_{rot} is low.

Nevertheless, to have more confidence about the thesis methodology, an additional parameter estimation is performed, this time using the structural GP and the log joint probability GP[13]

---

[12] Multiplying 20 walkers per 15,000 steps, and considering around 25 minute per finite element model evaluation, gives 125,000 computing hours.

[13] Given that the bridge model predictions and the log probabilities are obtained with cheap-to-compute Gaussian processes, this parameter estimation is performed in just a few minutes.

combined. Convergence is reached in 10 steps, meaning that only 40 structural GP calls are made. The posterior values presented in Table 24 show a remarkable resemblance with Table 23, which demonstrates that a very high accuracy with respect to a direct MCMC sampling is possible with the methodology chosen for this thesis, even when using noisy real measurements. The similarity is also observed in the posterior density plots from Figure 39.

Table 23: Posterior values for Sub-case 1 - Physical GP.

Parameter	Median	MAD
E_c (GPa)	27.30	0.54
K_rot (Nmm/rad)	7.94 E+11	1.80 E+11

Table 24: Posterior values for Sub-case 1 – Log joint probability GP + Structural GP.

Parameter	Median	MAD
E_c (GPa)	27.30	0.55
K_rot (Nmm/rad)	7.96 E+11	1.73 E+11

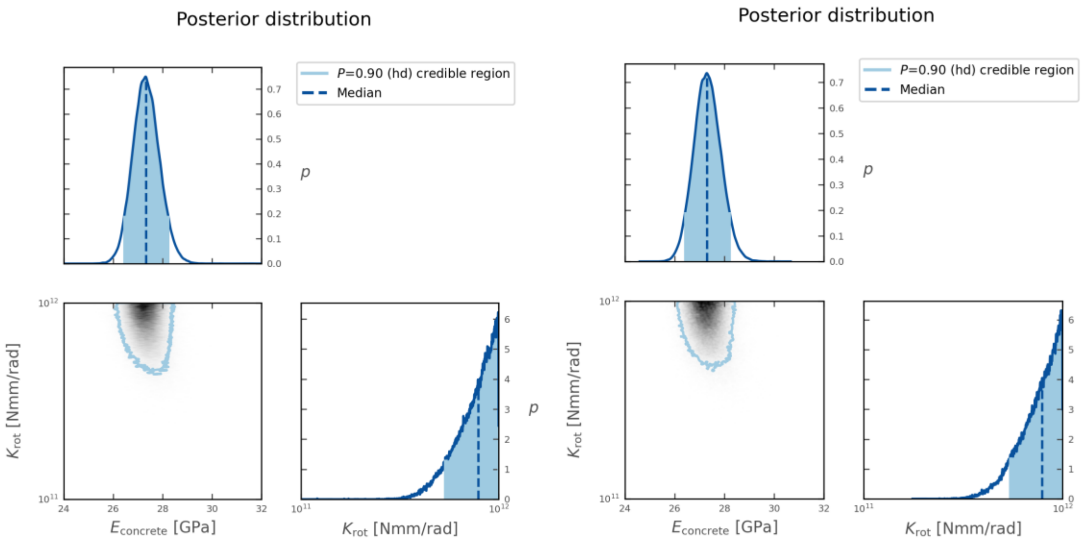


Figure 39: Zoomed-in posterior distribution for Sub-case 1 obtained using a multi-output GP surrogate to predict structural responses. The left plot shows was built by directly sampling with MCMC while the right one used an additional GP surrogate for the log joint probability. Diagonal: 1D marginal densities. Off diagonal: 2D contour plot of joint probability density.

The differences in results between the parameter estimations performed with and without the structural GP surrogate may come from discrepancies in predictions with respect to the ones obtained from DIANA FEA, as is illustrated in Figure 40. Notice that even though the R^2 scores are close to 1.0, equation (10) squares the error to obtain the log likelihood, amplifying the discrepancy in the posterior distributions.

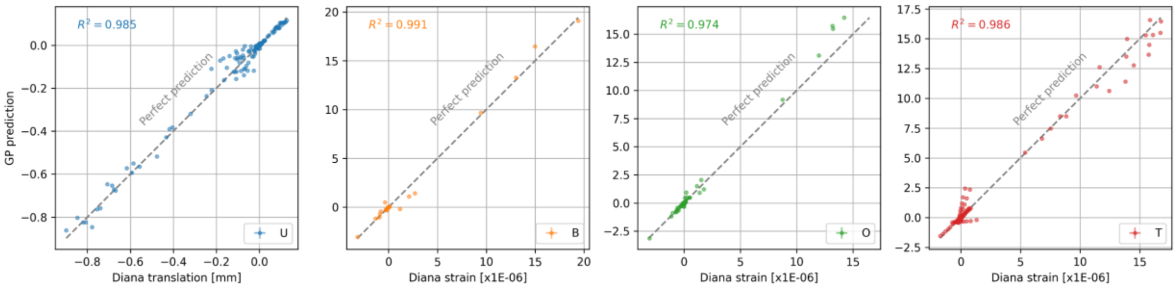


Figure 40: Comparison of a priori physical GP predictions (y-axis) vs Diana model predictions (x-axis) for strain ('B', 'O' and 'T') and translation ('U') sensor groups.

The convergence criterion from Section 3.5.3 is used. Specifically, convergence is reached when the variation of the median with respect to its median absolute deviation is below the threshold $\epsilon = 0.1$ in three consecutive steps for both parameters. For sub-case 1 this occurs in Step 9, meaning that only 36 additional finite element model evaluations are necessary to obtain an accurate posterior distribution, showing the remarkable efficacy of the active learning posterior methodology. In order to observe the tendencies, the analysis is continued until Step 15. Figure 41 shows the evolution of the estimated parameters. The medians and MADs remain flat after convergence is reached.

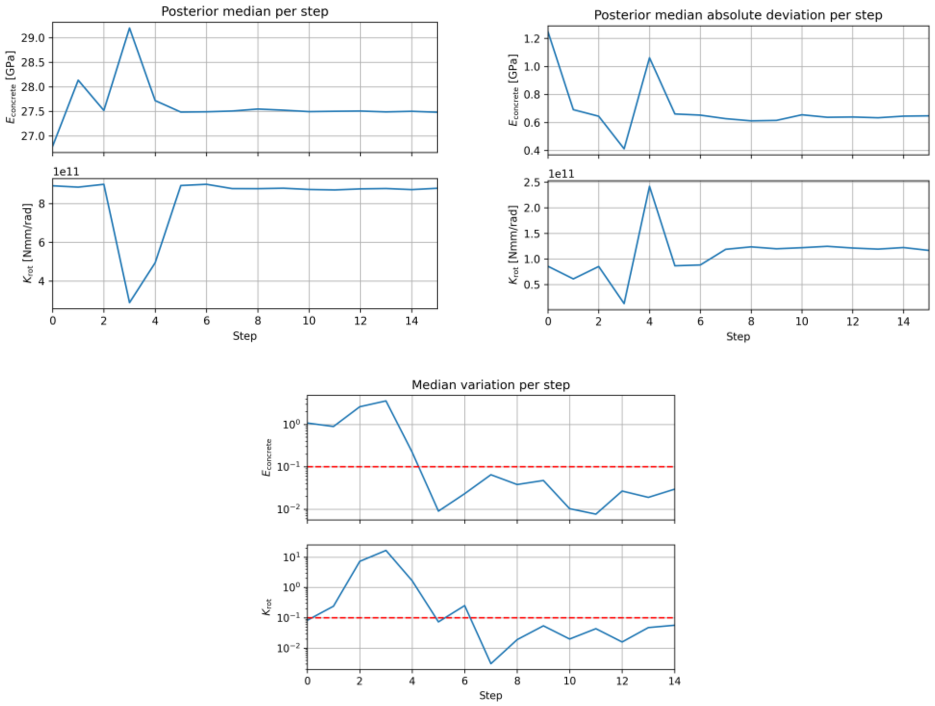


Figure 41: Evolution of MCMC parameter estimation per step for Sub-case 1, with four new samples drawn at each of them. Step 0 only uses the initial samples, i.e., before the active learning. Top left: Median values. Top right: Maximum absolute deviations. Bottom: Variation of the mean with respect to previous step normalized by the MAD. The red dashed line represents the convergence threshold set for this case study: 0.10.

Figure 42 shows the locations of the new points drawn by the active learning algorithm. These new samples are concentrated around the high density areas in E_c. In contrast to the synthetic case study, a distance constraint has been imposed in order to avoid points too close to each other, which explains the even distribution of orange dots. The fact that the points are aligned vertically demonstrates that the likelihood is less sensitive to K_{rot} for values in the order of 10^{12} Nmm/rad.

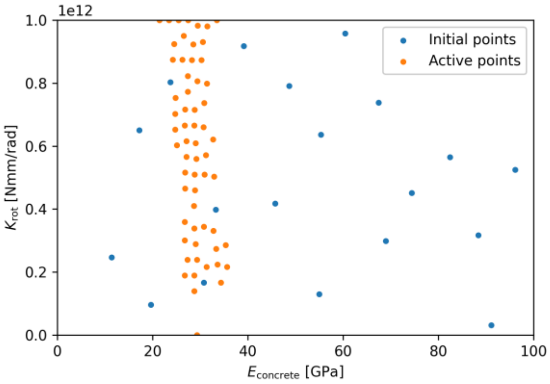


Figure 42: Samples used to surrogate the log joint probability GP surrogate for Sub-case 1. The blue dots are the initial Latin hypercube points while the orange are actively drawn.

The results obtained in Sub-case 1 serve to validate[14] the methodology proposed in this thesis. Some additional results related to the agreement between the a posteriori model predictions and measurements are presented in Appendix C.

6.3 Sub-case 2

Sub-case 2 uses only the strain sensors located in the path of the optic fibre in order to make a fair comparison with Sub-case 3. The emphasis in this section is on the a posteriori model predictions, therefore, the results about convergence of the algorithm and the distribution of drawn samples are not shown here but in Appendix D. Since the posterior active learning methodology was validated in the previous section, the structural GP surrogate is no longer used.

The settings for the posterior active learning and the MCMC are the same as for Sub-case 1. Table 25 shows the posterior median and MAD for the converged state. As it happened for the previous case, the value of the rotational stiffness K_{rot} is significantly increased with respect to the prior model. The median value of the concrete elastic modulus E_c is 31% lower than the median for Sub-case 1 and 51% lower than the a-priori value.

Table 25: Posterior values for Sub-case 2.

Parameter	Median	MAD
E_{c} (GPa)	18.92	1.20
K_{rot} (Nmm/rad)	7.90 E+11	1.79 E+11

The posterior probability densities are presented in Figure 43. The extent of the 0.90 highest density credible region is greater than in Sub-case 1, which is expected since using only a fraction of the sensors would naturally imply a higher uncertainty in the model predictions;

---

[14] Strictly speaking, a full order MCMC would be necessary to fully validate the methodology

however, the posterior median of E_c from Sub-case 1 is outside this credible interval. In other words, the parameter estimation of Sub-case 2 not only gives bad estimates for medians but also predicts an unrealistically narrow credible region.

A possible explanation for observed differences with Sub-case 1 is that this case mostly relies on sensor from group 'T', which showed a high number of predictions outside the 0.99 credible region of the measurements even for Sub-case 1 (see Appendix C). Rózsás, et al. (2022) performed a parameter estimation with all discrete strain sensors and the agreement with Sub-case 1 was much better, which would back up the idea that group 'T' sensors might not be too reliable.

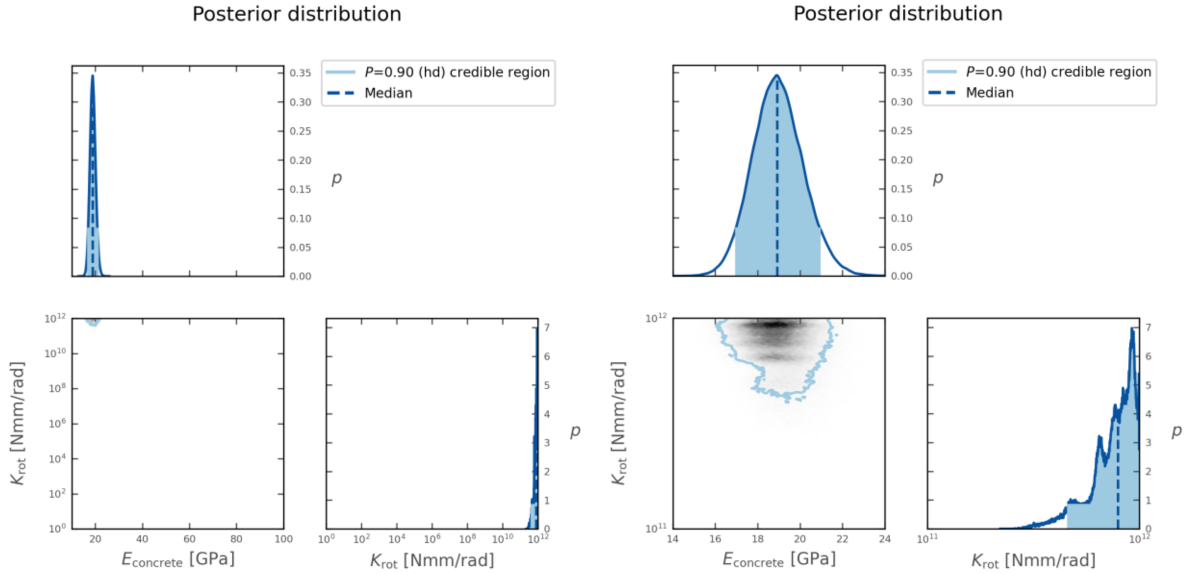


Figure 43: Converged posterior distribution for Sub-case 2. The left plot shows the same axis limits as the prior while the right one presents a zoomed-in view. Diagonal: 1D marginal densities. Off diagonal: 2D contour plot of joint probability density.

The posterior median values from Table 25 are used to obtain finite element model predictions which are then compared to the sensor measurements in Figure 44. The R^2 scores for the four sensor groups are considerably worse the a priori model. Striking is that the group that contains 80% of the sensor used for this case, sensor group 'T', is the one with the worst fit, although this might be explained if we accept the hypothesis that this sensor group is not reliable.

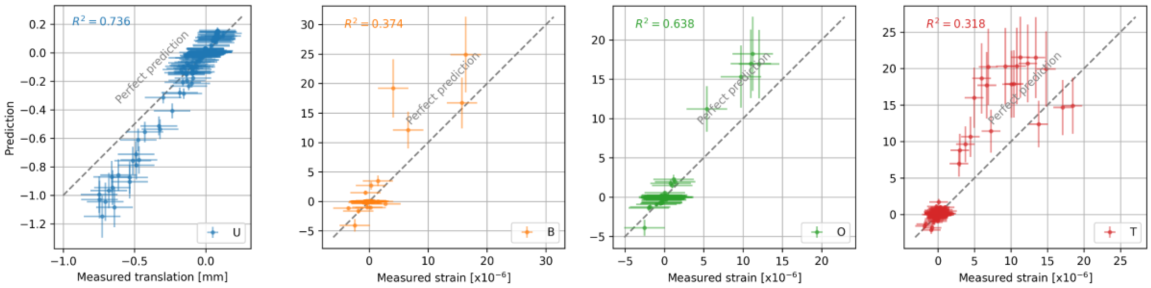


Figure 44: Comparison of model predictions obtained with Sub-case 2 converged posterior median values (y-axis) vs measurements (x-axis) for translation ('U') and strain ('B', 'O' and 'T') sensor groups. The vertical and horizontal lines on the dots indicate the 0.99 highest density credible intervals for model prediction uncertainty and measurement uncertainty, respectively.

Figure 45 shows strain diagrams for the bottom fibre of the deck at the centreline of Truck 1 lane. There is not only a significant discrepancy between predictions and measurements, but the 0.99 measurement credible intervals mostly do not cross the model prediction credible interval in any of the peak areas.

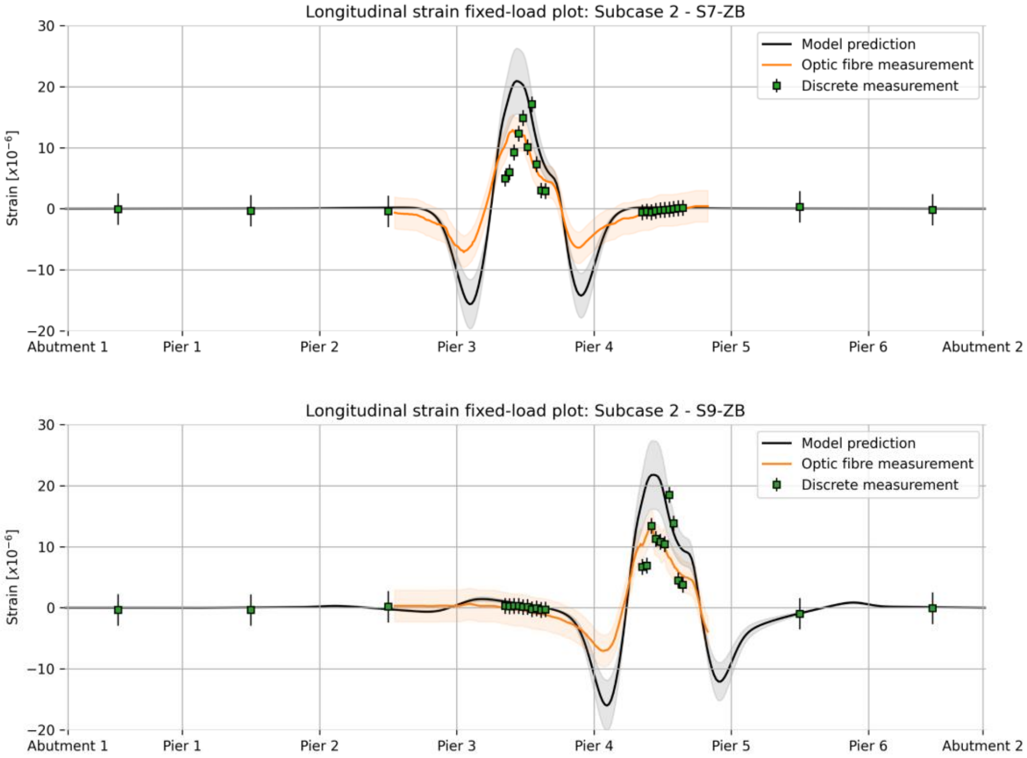
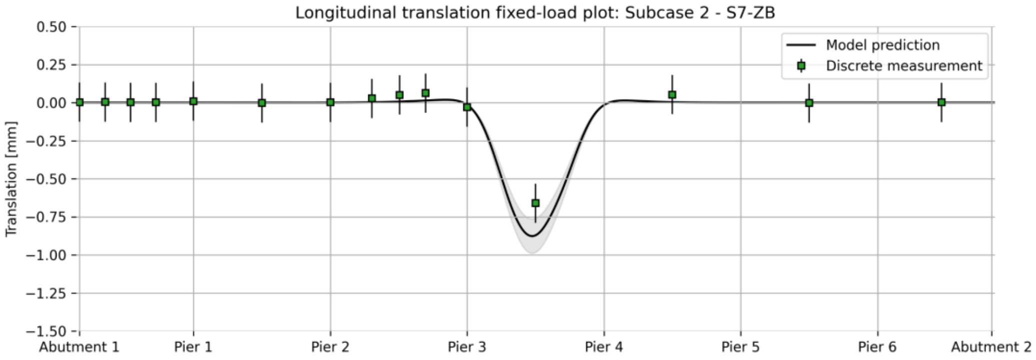


Figure 45: Longitudinal strain diagrams at centreline of Truck 1 lane (deck bottom fibre). The model predictions are obtained with the posterior median parameters from Sub-case 2. The grey-shaded region indicate the 0.99 highest density credible intervals for the model prediction uncertainty. The orange-shaded region and black vertical lines indicate the 0.99 highest density credible intervals for measurement uncertainty for optic fibre and discrete measurements, respectively. Top: load case S7-ZB. Bottom: load case S9-ZB.

Also the vertical translation diagrams in Figure 46 display a considerable disagreement for the sensors below the load position, with a very small overlap of the credible intervals for load case S7-ZB and no overlap at all for S9-ZB.

Figure 47 shows a comparison of the strain predictions and measurements of the optic fibre, hence evaluating the prediction accuracy for unseen data. The R^2 score here is worse than the a priori model as well. It can also be observed that, in terms of absolute values, the predictions are generally higher than the measurements, i.e., rotated counter clockwise from the diagonal. This confirms that the actual stiffness of the deck, represented by E_c, is higher than what was obtained in for this sub-case.



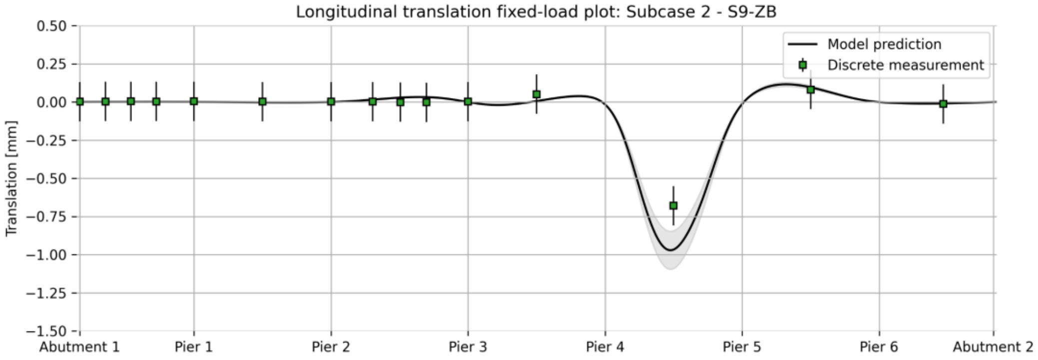


Figure 46: Vertical translation diagrams at centreline of Truck 1 lane. The model predictions are obtained with the posterior median parameters from Sub-case 2. The grey-shaded region indicate the 0.99 highest density credible intervals for the model prediction uncertainty. The orange-shaded region and black vertical lines indicate the 0.99 highest density credible intervals for measurement uncertainty for optic fibre and discrete measurements, respectively. Top: load case S7-ZB. Bottom: load case S9-ZB.

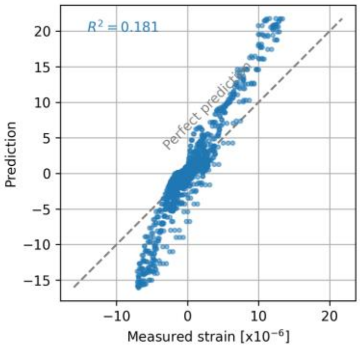


Figure 47: Comparison of model predictions obtained with Sub-case 2 converged posterior median values (y-axis) vs measurements (x-axis) for the optic fibre sensor

6.4 Sub-case 3

Sub-case 3 uses the transversal and bottom legs of the optic fibre for the parameter estimation as shown in Figure 27 and Table 13, in order to compare the information obtained by using high resolution optic fibre measurements and discrete strain measurements. Results about convergence and the distribution of drawn samples are shown in Appendix E.

The posterior median values and MADs for the converged state are presented in Table 26. As it happened with the previous sub-cases, the inferred rotational stiffness K_{rot} is orders of magnitude above the a priori value; however the median is still around 48% and 42% lower than what was obtained for Sub-cases 1 and 2, respectively. On the other hand, the concrete elastic modulus E_c is closer to the a priori model than the previous cases, with the median only around 11% lower than its initially assumed value.

Table 26: Posterior values for Sub-case 3.

Parameter	Median	MAD
E_{c} (GPa)	34.27	1.14
K_{rot} (Nmm/rad)	4.59 E+11	1.29 E+11

The converged posterior probability densities are presented in Figure 48. The inferred 0.90 highest density credible region is larger than in Sub-case 1, but smaller than Sub-case 2. Also, the shape of the 1D marginal distribution over K_{rot} is no longer tilted towards its right boundary, which was an undesired outcome from the marginals in Sub-case 1 and 2.

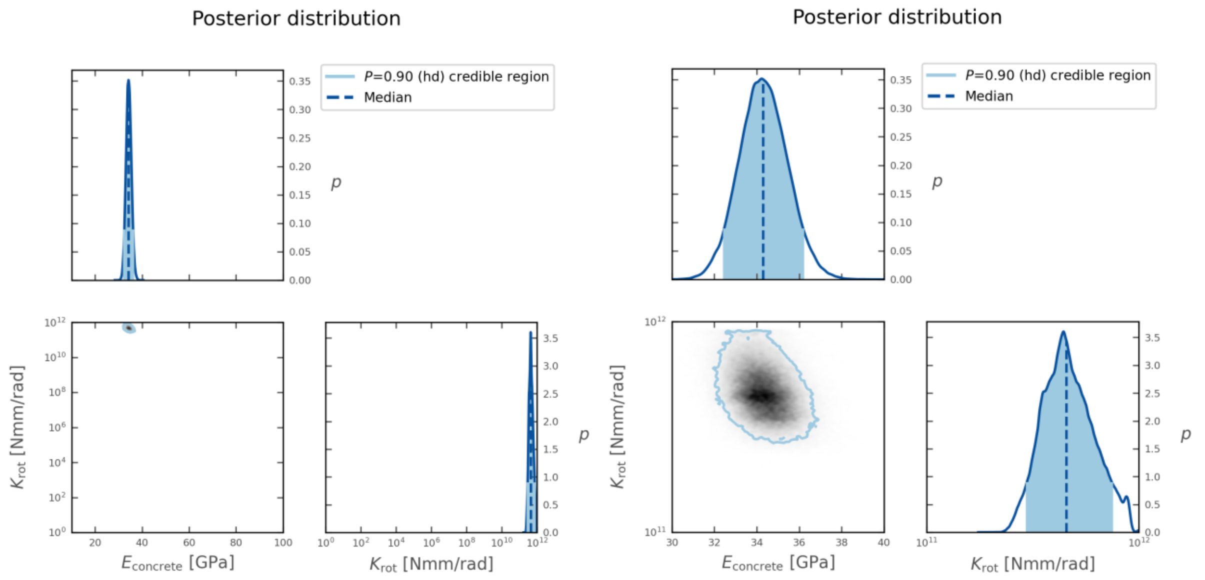


Figure 48: Converged posterior distribution for Sub-case 3, with four new samples drawn at each of them. Step 0 only uses the initial samples, i.e., before the active learning. Top left: Median values. Top right: Maximum absolute deviations. Bottom: Variation of the mean with respect to previous step normalized by the MAD. The red dotted line represents the convergence threshold set for this case study: 0.10.

The posterior median values from Table 26 are used to obtain finite element model predictions which are then compared to the sensor measurements in Figure 49. For all sensor groups, an improvement on the R^2 scores is obtained in Sub-case 3 with respect to the a priori model. Even more, for sensor groups ‘U’ and ‘O’, the credible intervals for all points cross the dotted diagonal line.

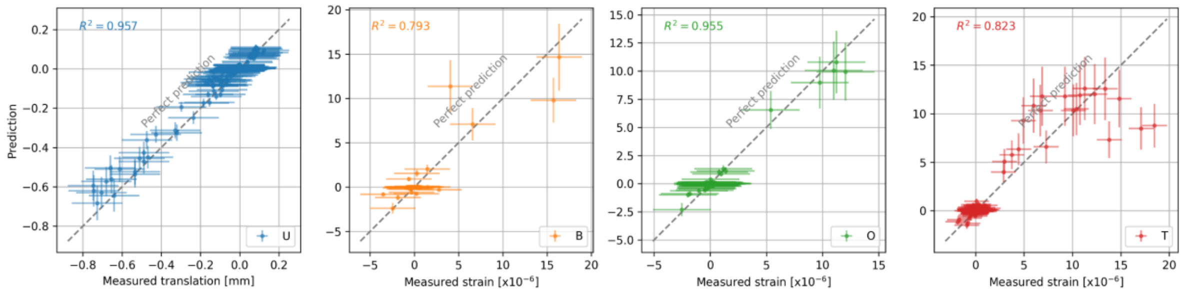


Figure 49: Comparison of model predictions obtained with Sub-case 3 converged posterior median values (y-axis) vs measurements (x-axis) for translation ('U') and strain ('B', 'O' and 'T') sensor groups. The vertical and horizontal lines on the dots indicate the 0.99 highest density credible intervals for model prediction uncertainty and measurement uncertainty, respectively.

On the other hand, both sensor groups 'B' and 'T' contain points whose credible intervals do not reach the diagonal. The predictions for the optic fibre from Figure 50 show the best agreement with the measurements across all sub-cases, with an R^2 score 3% and 11% greater than the prior model and Sub-case 1, respectively.

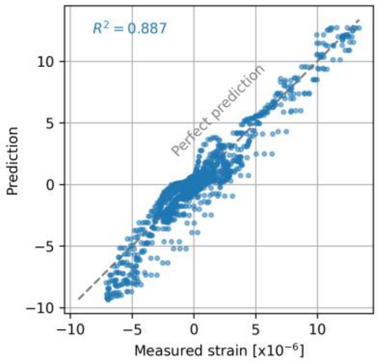
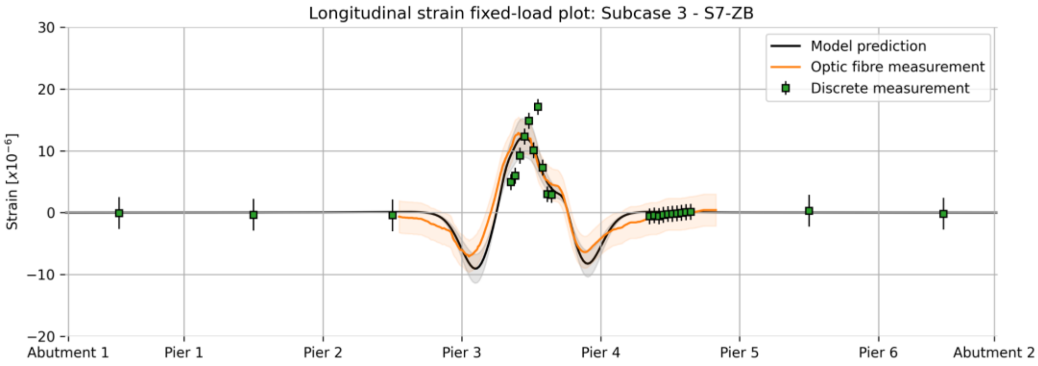


Figure 50: Comparison of model predictions obtained with Sub-case 3 converged posterior median values (y-axis) vs measurements (x-axis) for the optic fibre sensor.

In summary, it could be said that after the structural parameter estimation of Sub-case 3 the model predictions for all sensor groups are improved with respect to the prior model. When comparing against Sub-case 1, the predictions are generally improved for all strain sensor groups, while their match is slightly lower for the translation sensors, which was expected since 'U' sensors were not used for Sub-case 3.

The previous conclusion can be visually explained by looking at the strain plots in Figure 51. The match on the peak areas between the model prediction and the optic fibre measurements is almost perfect, while some slight differences appear near the piers, but still within the credible regions. The discrete strain measurements from group 'T' still show considerable differences with both the optic fibre measurements and the predictions. At this point, it is clear that this sensor group contains issues that may steer Sub-cases 1 and 2 towards wrong estimations.



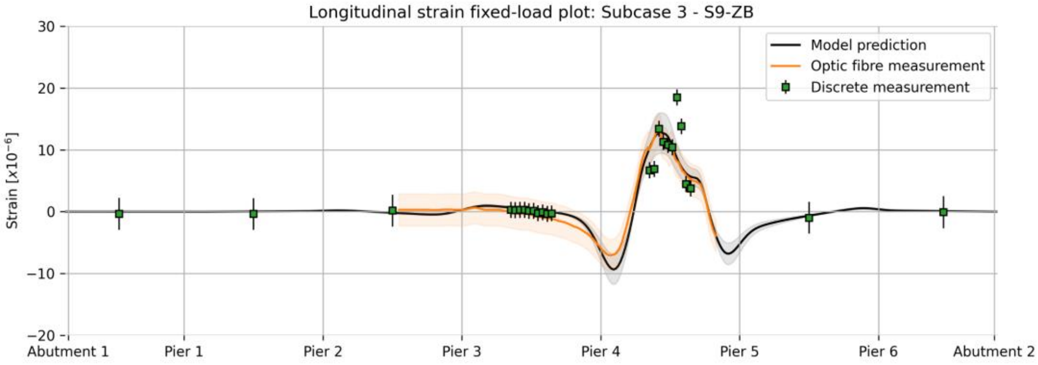


Figure 51: Longitudinal strain diagrams at centreline of Truck 1 lane. The model predictions are obtained with the posterior median parameters from Sub-case 3. The grey-shaded region indicate the 0.99 highest density credible intervals for the model prediction uncertainty. The orange-shaded region and black vertical lines indicate the 0.99 highest density credible intervals for measurement uncertainty for optic fibre and discrete measurements, respectively. Top: load case S7-ZB. Bottom: load case S9-ZB.

Finally, Figure 52 confirms that the translation predictions are improved with respect to the a priori model but are less accurate than the parameter estimation in Sub-case 1 that used the translation sensors. Nevertheless, for all sensors in load cases S7-ZB and S9-ZB the measurement and model prediction credible intervals overlap.

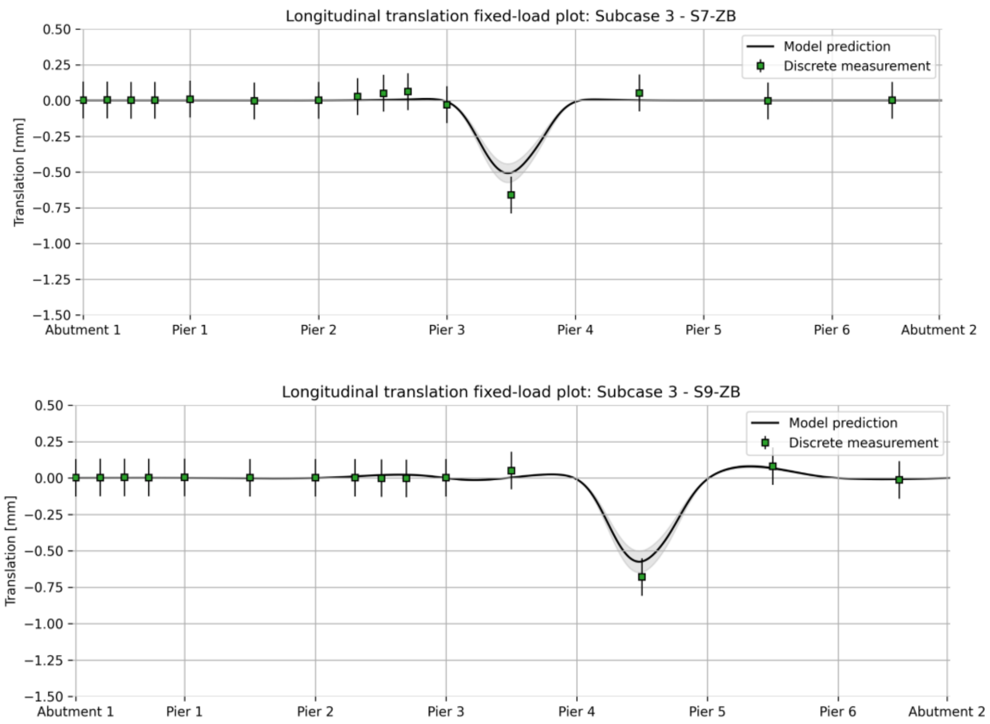


Figure 52: Vertical translation diagrams at centreline of Truck 1 lane. The model predictions are obtained with the posterior median parameters from Sub-case 3. The grey-shaded region indicate the 0.99 highest density credible intervals for the model prediction uncertainty. The orange-shaded region and black vertical lines indicate the 0.99 highest density credible intervals for measurement uncertainty for optic fibre and discrete measurements, respectively. Top: load case S7-ZB. Bottom: load case S9-ZB.

6.5 Sub-case 4

All parameter estimation cases performed so far have been focused on inferring structural parameters, while using fixed statistical parameters to describe the model and measurement uncertainties. As discussed in 5.4, the statistical parameters for this case study were estimated based on expert judgement, and thus, are not free of uncertainty. In order to assess their influence in the parameter estimation, Sub-case 4 aims to identify the statistical parameters σ_{meas}, cov_{model} and L_{corr}, along with the physical parameters E_{c} and K_{rot}, resulting on a five-dimensional parameter estimation case.

The initial dataset of 840 samples (see Section 5.4) and their respective log joint probability output is split so that 75% is used to start the posterior active learning procedure, while the remaining is employed to test the accuracy of the GP surrogate[15]. For this sub-case, 10 new samples are drawn per step based on the exponentiated variance utility function. The algorithm was ran for 9 steps without reaching convergence, therefore, it is not possible to make general conclusions about the influence of the statistical parameters on Bayesian parameter estimation; nevertheless, the posterior distribution estimated after the final step is shown for completeness. The MCMC sampling was performed using 50 walkers and 15,000 final steps.

---

[15] The GP surrogate is trained with an anisotropic *Matern* kernel with $\nu = 1.5$

In general, the posterior from Figure 53 is much narrower than the prior distribution shown in Figure 32, with very low median absolute deviations (MAD) values for all parameters, (see Table 27). This low degree of uncertainty might not be realistic considering that convergence was not achieved. Still, some resemblance with Sub-case 3 is found for the physical parameters, with E_{c} having a posterior median slightly lower than the a priori model, and K_{rot} being much higher than its initially assumed value. In contrast, very different marginal 1D distribution shapes are obtained for these two parameters with respect to sub-case 3.

Table 27: Posterior values for Sub-case 4.

Parameter	Median	MAD
E_{c} (GPa)	31.50	0.85
K_{rot} (Nmm/rad)	9.95 E+11	3.91E+09
σ_{meas}	0.33	0.004
cov_{model}	0.32	0.001
L_{corr} (m)	3.88	0.02

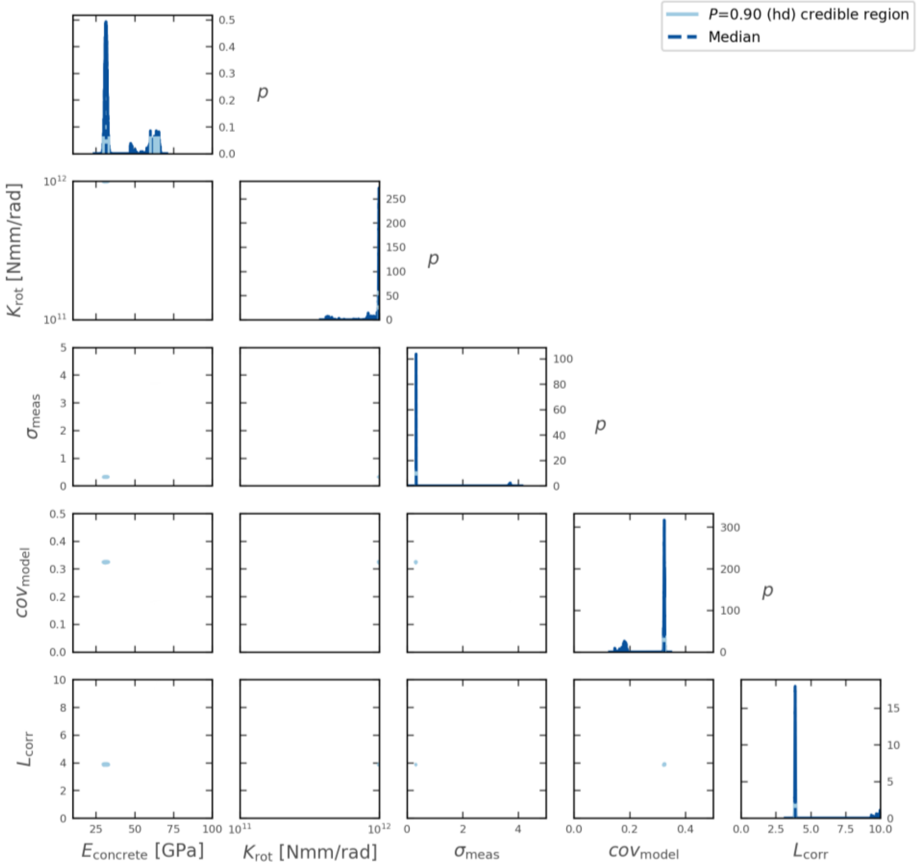


Figure 53: Posterior distribution for Sub-case 4. Diagonal: 1D marginal densities. Off diagonal: 2D contour plot of joint probability density.

The remaining of this section is dedicated to aspects that might explain the unsuccessful parameter estimation. Figure 54 shows a comparison between the real log joint probability values and the GP predictions for the 25% test dataset. The left plot shows a very poor global approximation, with a negative R^2 score, mainly driven by inaccurate predictions on low probability regions (above 1E+4). Indeed, the real values grow beyond 1E+7 while the GP predictions remain 'flat' at around 2E+4. On the other hand, the right plot shows that the GP predictions are far better for the high probability values, i.e., below 1E+4.

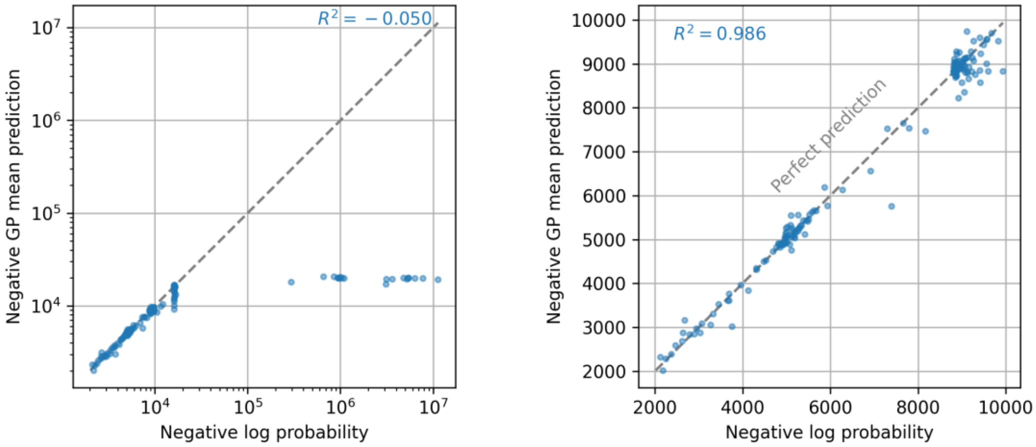


Figure 54: Comparison of negative log joint probability (x-axis) vs negative GP surrogate predictions (y-axis). The left plot is in logarithmic scale and shows the entire dataset. The right plot is a linear scale zoomed view to values below 1E+4.

In theory, a surrogate with such a high R^2 score value in high probability areas should be accurate enough for parameter estimation; however, it should be noted that the log joint probability varies faster across statistical parameters than physical parameters for this case. In other words, for a given combination of statistical parameters, only small variations of the log probability (in percentage) occur in the physical parameter domain. In order to evaluate if the GP surrogate is capable of representing this behaviour, a few samples from the test set are selected.

Table 28: Comparison of real log joint probability values vs GP surrogate predictions for different samples, all of them with $\sigma_{\text{meas}} = 0.70$ and $cov_{\text{model}} = 0.10$. The real log probabilities are displayed in a descending order.

E_c (GPa)	K_{rot} (Nmm/rad)	L_{corr} (m)	Real log probability	Real order	GP prediction	GP order	GP / Real ratio
33.2	3.98E+11	1.5	-5028.1	1	-5289.7	7	1.05
33.2	3.98E+11	0.5	-5106.3	2	-5550.7	8	1.09
39.1	9.18E+11	0.5	-5121.6	3	-5262.2	6	1.03
45.7	4.18E+11	1.5	-5136.1	4	-5209.4	5	1.01
17.2	6.50E+11	3.0	-5180.3	5	-5069.3	2	0.98
30.7	1.66E+11	0.5	-5187.5	6	-5037.6	1	0.97
48.7	7.91E+11	3.0	-5209.9	7	-5160.9	3	0.99
11.4	2.47E+11	3.0	-5420.6	8	-5119.6	4	0.94

Table 28 shows a comparison of the real log joint probability and the GP predictions for the selected samples. The last column presents ratios between predictions and real values very close to 1.0, with differences below 10%; however, these discrepancies are enough to completely alter the order of the predictions, as seen in the penultimate column of the table. Considering that MCMC methods are not directly driven by the log probabilities of the considered samples but by their difference[16], it could be inferred that the GP surrogate is not accurate enough. Indeed, the GP prediction errors from Table 28 are not negligible when compared to the difference of real log joint probabilities within the samples.

It is also interesting to see how effective the posterior active learning algorithm was on drawing samples on the high probability regions. Figure 55 shows the real log joint probability for the 90 drawn samples. It can be noted that high probability values were not consistently selected by the algorithm, with the exception of steps 6 and 7. It could be the case that more steps were necessary for the algorithm to select 'better' samples in a more stable way.

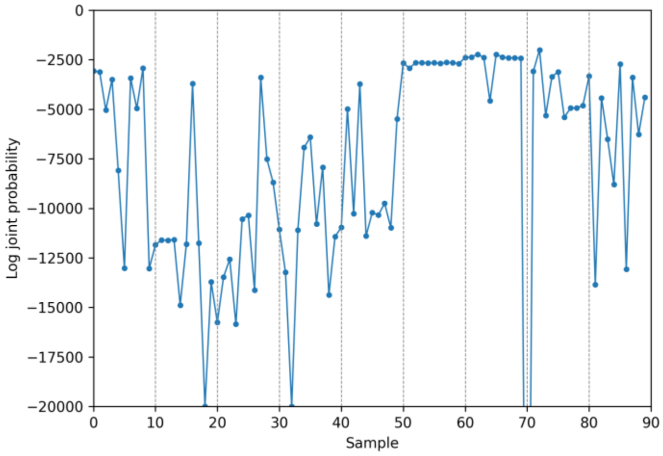


Figure 55: Real log joint probability of actively drawn samples. The vertical grey dashed lines mark the beginning of a new step of the algorithm.

6.6 Discussion

6.6.1 Performance

One of the main goals of this thesis is to reduce the computation time of Bayesian system identification of real-world civil engineering structures. On that line, the total running times for the first three sub-cases of Bridge 705 are projected under the four different scenarios described below, including the formulas to calculate them:

- *Full order MCMC*: default scenario where the finite element model is evaluated at each MCMC sample.

$$t_{\text{projected}} = n_{\text{MCMC}}(t_{\text{likelihood}} + t_{\text{FEA}}) \tag{38}$$

---

[16] As explained in Section 3.4, the acceptance of a proposed sample in MCMC depends on the division of its probability with the probability of the current sample. This is equivalent to applying the exponential function to the difference of log probabilities.

- *Structural GP*: scenario where the finite element model responses are approximated with a GP surrogate trained with a grid of samples, emulating Rózsás, et al. (2022).

$$t_{\text{projected}} = n_{\text{grid}} \cdot t_{\text{FEA}} + n_{\text{MCMC}}\left(t_{\text{likelihood}} + t_{\text{physGP}}\right) \tag{39}$$

- *Log joint probability GP*: scenario that uses the methodology proposed in this thesis, i.e., an actively built GP surrogate of the log joint probability density.

$$t_{\text{projected}} = (n_{\text{latin}} + n_{\text{active}}) \cdot (t_{\text{likelihood}} + t_{\text{FEA}}) + n_{\text{MCMC}} \cdot t_{\text{logprobGP}} \tag{40}$$

- *Log joint probability GP with cloud-based parallelization*: similar to the previous scenario but adding cloud-based parallelization.

$$\begin{aligned} t_{\text{projected}} = {} & (n_{\text{latin}} + n_{\text{active}}) \cdot (t_{\text{likelihood}} + t_{\text{FEA}} + t_{\text{overhead}})/n_{\text{parallel}} \\ & + n_{\text{MCMC}} \cdot t_{\text{logprobGP}} \end{aligned} \tag{41}$$

where:

- n_{MCMC} is the total number of MCMC samples
- n_{grid} is the number of grid points used to fit the structural GP
- n_{latin} is the number of initially drawn Latin hypercube samples
- n_{active} is the number of actively drawn samples until reaching convergence
- n_{parallel} is the number of parallel FE model evaluations
- $t_{\text{likelihood}}$ is the computation time for a single likelihood evaluation
- t_{FEA} is the computation time for a single DIANA FEA model evaluation
- t_{physGP} is the computation time for a single structural GP prediction
- $t_{\text{logprobGP}}$ is the computation time for a single log joint probability density prediction
- t_{overhead} is the overhead time due to cloud-based parallelization

Table 29 shows the number of evaluations per component.

Table 29: Number of evaluations of Bayesian workflow components per sub-case

Case	n_{MCMC}[17]	n_{grid}	n_{latin}	n_{active}	n_{parallel}
Sub-case 1	340,000	112	20	36	4
Sub-case 2	340,000	112	20	60	4
Sub-case 3	340,000	112	20	36	4

Table 30 presents the unitary computation times estimated by performing several test runs and calculating their average time. A laptop with 16 GB of RAM and a 1.7 GHz CPU with 8 cores was used.

---

[17] Obtained by multiplying the number of walkers (20) with the total number of MCMC steps (17,000)

Table 30: Unitary computation times of Bayesian workflow components per sub-case.

Case	t_{FEA}	t_{overhead}	$t_{\mathrm{likelihood}}$[18]	t_{physGP}[19]	$t_{\mathrm{logprobGP}}$[20]
Sub-case 1	1500 s	600 s	0.09 s	0.34 ms	83 µs
Sub-case 2	1500 s	600 s	0.02 s	0.32 ms	86 µs
Sub-case 3	1500 s	600 s	9.18 s	0.76 ms	83 µs

Figure 56 presents the projection of the total running time for the four scenarios using the previous information. The default case where no GP approximation is used for the parameter estimation (blue bar) gives total runtimes in the realm of years for the three sub-cases, making it unfeasible for practical purposes, i.e., without super computers. On the other hand, using a physical GP (orange bar) gives reasonable running times for cases with discrete sensor measurements (Sub-cases 1 and 2); however, the method becomes unmanageable for the case with optic fibre measurements (Sub-case 3) where covariance matrices with thousands of rows and columns have to be evaluated. This is where the methodology proposed on this thesis (green and red bars) shows a clear advantage since the log likelihood is not directly computed but approximated with a log probability GP. It can be seen that in all sub-cases the Bayesian parameter estimation is computed within a few hours, showing a remarkable efficiency when comparing it with the other approaches.

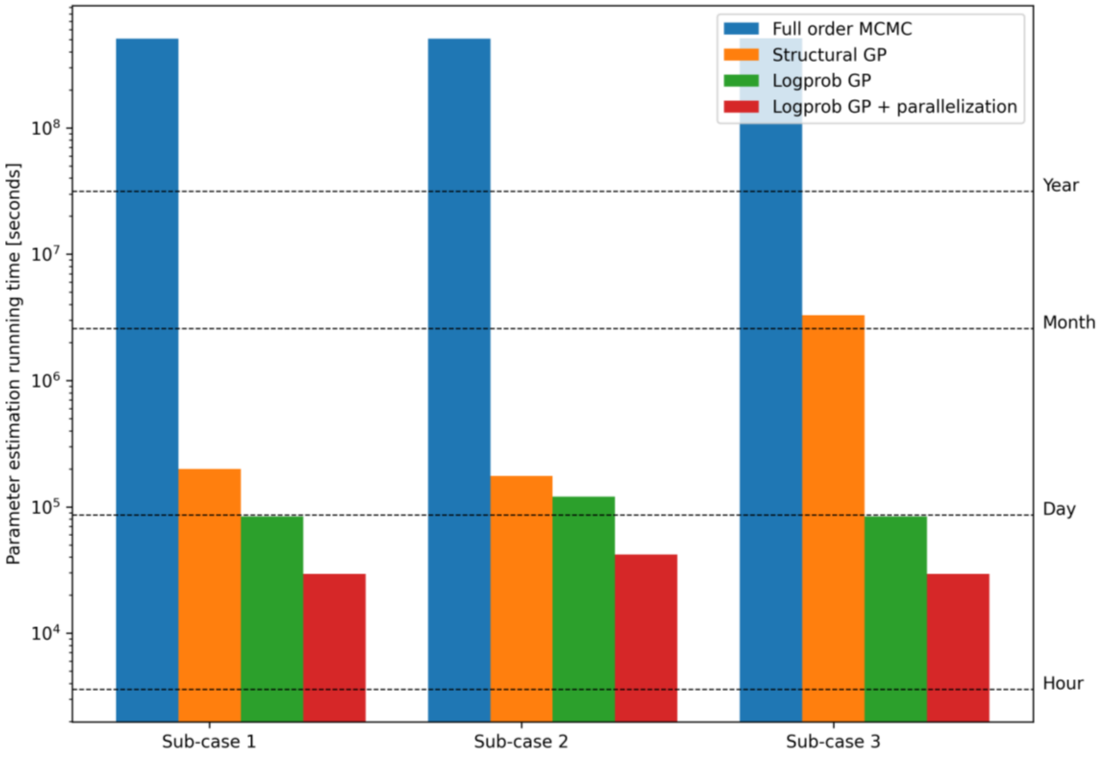


Figure 56: Estimation of total Bayesian parameter estimation runtime for each sub-case. The training times for the Gaussian process are not included as they are negligible compared to the total times

---

[18] Differences are due to the number of measurements included on each sub-case

[19] Differences are due to the number of predictions per GP call for each sub-case

[20] Differences are due to the number of samples used to fit the converged GP for each sub-case

The comparison above is based on a two-dimensional problem. With increasing dimensionality, the running times per approach are expected to scale differently. For example, the structural GP scenario requires a global surrogate approach (i.e., accurate in the entire parameter domain) which generally scales poorly with the number of dimensions. In contrast, the log joint probability GP approach proposed in this thesis aims to only select points that contribute the most to the accuracy of the posterior distribution, which intuitively would demand a lower number of finite element model evaluations.

Nevertheless, building *global* surrogates of the finite element responses has also a few advantages. For example, they may be easier to approximate with Gaussian processes than log probabilities, which can be highly nonlinear and fast-varying (Wang & Li, 2017). In addition to that, while the structural GP can be reused for different parameter estimations, the log probability GP has to be obtained for every new case. Furthermore, a structural GP can be used with other Bayesian inference methods such as Nested Sampling, as shown by Chai (2019), while the log probability GP specifically works with MCMC methods. In general, different conditions could make one surrogating approach more desirable than the other; however, for this case study, it has been proven that the proposed methodology is both accurate and computationally effective.

It is also interesting to evaluate the monetary cost of the cloud-based parallel finite element model evaluations used in this chapter. As mentioned in Section 3.6, the parallelization scheme relies on *Microsoft Azure*, which charges the user based on the amount of computational resources used per unit of time.

The total costs per case study are shown in Table 31. The hourly cost of the selected virtual computers (properties shown in Section 5.5) is 0.0875 USD[21], and the duration of each parallel step was estimated in 0.5 hours, including the FEA running time and cloud overhead time. The initial Latin hypercube samples are accounted separately since they were ran only once.

Table 31: Total cost of finite element parallel runs per case study.

Description	Steps	Runs per step	Step duration	Unit cost	Total cost
Initial sampling	1	20	0.5 hours	0.0875 USD	0.875 USD
Sub-case 1	9	4	0.5 hours	0.0875 USD	1.575 USD
Sub-case 2	15	4	0.5 hours	0.0875 USD	2.625 USD
Sub-case 3	9	4	0.5 hours	0.0875 USD	1.575 USD

The total costs are remarkably low and show a great potential for the application of cloud-based parallelization, especially considering that prices will diminish over time with the progress of the technology. Nevertheless, its use in practical situations will depend on users having access to an IT infrastructure that makes it possible. At the moment, cloud-based parallelization is not a common feature in commercial finite element modelling software,

---

[21] Price consulted in https://azureprice.net/?tier=spot&cpu=14,416 on July 28th, 2022

although things might be moving in that direction, as shown by the recent introduction of a cloud environment for the finite element software *Abaqus* (Dassault Systemes, 2022).

6.6.2 Utility of optic fibre measurements for Bayesian inference

In order to assess the benefit of the optic fibre measurements for the identification of structural parameters, the median values obtained with the parameter estimations were used to make finite element predictions, which were then compared against the discrete sensor measurements predictions. Table 32 shows a summary of the coefficients of determinations (R^2 score) presented in the previous sections.

Table 32: Summary of R^2 scores per sensor set and sub-case.

Sensor set	A priori	Sub-case 1	Sub-case 2	Sub-case 3
Translation	0.89	0.97	0.74	0.96
Discrete strain	0.84	0.82	0.44	0.86
Optic fibre	0.86	0.79	0.18	0.89

It is clear that the a priori model already gives good predictions, with all R^2 scores with values above 0.80. For Sub-case 1, the parameter estimation improves the predictive capacity for translations but worsens it for discrete and optic fibre strains. Sub-case 2, on the other hand, yields the worst results with lower scores across all sensor groups when comparing to the a priori model. Finally, all sensor set coefficients are improved with Sub-case 3, with a translation R^2 score almost as good as the one obtained in Sub-case 1, even though no translation sensors were used.

It is also interesting to evaluate the parameter estimation uncertainty obtained in each sub-case. Table 33 shows the extension of the 1D 0.90 highest density credible interval. The most noticeable aspect is that the intervals obtained with Sub-case 1 and Sub-case 3 do not overlap, which might indicate an underestimation of the uncertainty for at least one of them. Sub-case 2 yields the wider credible intervals, which is in line with using a lower amount of measurements; however, they are considerably shifted from the posterior values for the other sub-cases. As previously mentioned, the strain sensor group 'T' forms the majority of the used sensors in Sub-case 2, which show erratic behaviour and a poor agreement when compare with both strain predictions and optic fibre measurements. A higher measurement uncertainty standard deviation for this group could have given more realistic medians.

Table 33: Summary of 0.90 highest density credible intervals extension per sub-case.

Sensor set	Sub-case 1	Sub-case 2	Sub-case 3
E_c (GPa)	26.41 - 28.55	16.95 – 20.89	32.39 – 36.15
K_{rot} (Nmm/rad)	6.89 E+11 - 1.00 E+12	4.96 E+11 - 1.00 E+12	2.47 E+11 – 6.71 E+11

In general, the results obtained with optic fibre strain sensors were much better than when using discrete strain sensors, and even though this might be due to some problematic discrete sensors, the optic fibre measurements were not entirely flawless. Nevertheless, since the optic fibre information was recorded at such high spatial rate, it was easier to determine tendencies, compare with model predictions and apply the appropriate corrections, as described in Appendix B.

6.6.3 Surrogating the log likelihood

The log joint probabilities were successfully approximated with Gaussian processes for the first three sub-cases using an anisotropic Matern kernel ($\nu = 2.5$); however, in some cases new samples were selected close to the corners of the parameter domain, even though these areas had very low likelihoods.

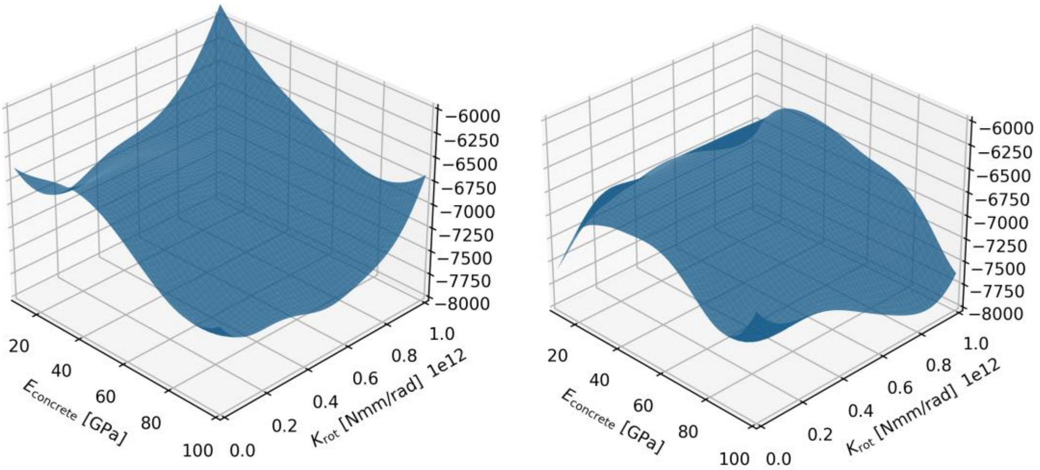


Figure 57: 3D view of the log probability GP mean predictions for Sub-case 3. Left: GP fitted using only the initial Latin hypercube samples. Right: GP fitted after reaching convergence.

Figure 57 shows the log probability GP predictions over the entire parameter domain for Sub-case 3. When the GP is fitted using only the initial training points, the mean predictions on the corners are significantly overestimated, as it can be appreciated on the left plot. The right plot show the GP after reaching convergence with exact FE model evaluations near the corners, as the algorithm selected points in these areas and found the actual values of the log probabilities. To avoid this issue, these corner points could be included in the initial sampling. Nevertheless, it should be noted that the algorithm corrected itself, although taking more steps to reach convergence.

On the other hand, the introduced parameter distance constraint proved effective to avoid the excessive clustering of new samples from Case Study 1 (see Section 4.5.2), hence allowing the selection of several unique new points to be run with cloud-based parallelization. Still, in some occasions all new points drawn for a single step appeared in the same location, separated only by the initially set minimal distances. In this situation, the gain in GP accuracy by drawing multiple points is not very different that doing it with just one. Furthermore, the distance constraints are expected to be highly case-specific. If one wants to guarantee that drawing a higher number of parallel points directly implies reaching convergence at a faster rate, this methodology should be refined.

Finally, the surrogation for Sub-case 4 is more complicated since the log likelihood of a point in the physical parameter domain can considerably change for different combinations of statistical parameters, as it was demonstrated in Section 6.5. As an example, Table 34 and Figure 58 show log joint probability values with different orders of magnitude. In this context, it is clear that the Gaussian process regression would quickly detect the log likelihood variations across statistical parameters but would struggle to do so for the physical parameters. Inferring statistical parameters under this methodology still requires more research.

Table 34: Log joint probability for different statistical parameters (E_c = 33.2 GPa, K_{rot} = 3.98E+11 Nmm/rad).

σ_{meas}	cov_{model}	L_{corr}	Log joint probability
0.01	0.01	10.00 m	- 4.29 E+06
0.01	0.50	0.01 m	- 8.20 E+05
5.00	0.50	10.00 m	-1.62 E+04
1.50	0.10	1.50 m	-8.83 E+03
0.30	0.40	3.00 m	-2.37 E+03

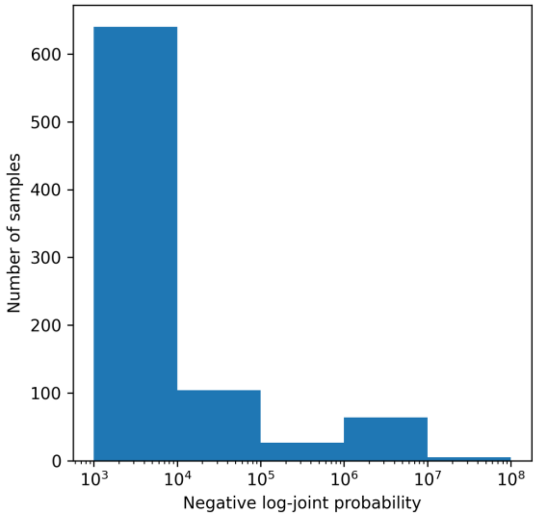


Figure 58: Logarithmic histogram of negative log joint probability for the initial samples of Sub-case 4.

7 Conclusion

7.1 Answer to research questions

With the results obtained in Chapters 4 and 6, the research sub-questions presented in Section 1.2 can be answered:

A. *How to include recent developments in surrogate modelling techniques and cloud-based parallelization in order to accelerate the Bayesian inference workflow?*

The main computational bottleneck of the Bayesian inference workflow for real-world civil engineering structures is usually the finite element model, whose evaluation is necessary to calculate the likelihood of the data. With this in mind, the methodology proposed in this thesis approximates the log joint probability with a Gaussian process surrogate, which is sequentially built by selecting new points on areas that are expected to highly contribute to the accuracy of the posterior. The number of finite element model evaluations needed under this approach is far lower than what is required for a directly applied MCMC method, with reductions by factors as high as 6000, as proven for both a synthetic (Sections 4.2 and 4.4) and a real case (Section 6.2), while obtaining almost equivalent posterior distributions. Furthermore, Section 6.6.1 demonstrates that predicting the log joint probability with a Gaussian process can be considerably faster than computing the exact log likelihood in cases with a high number of measurements, making the proposed method more attractive than strategies where the physical finite element responses are surrogated. Finally, the cloud-based parallelization strategy presented in Section 3.6 was successfully used in order to reduce the parameter estimation running time at a minimal monetary cost (see Section 6.6.1). The maximum number of simultaneous Diana model evaluations used was 10; however, its limit is only determined by the number of available software licenses.

B. *What gain in information content can be obtained from optic fibre strain sensors compared to conventional discrete sensors for the identification of structural model parameters?*

The large amount of data produced by distributed optic fibre sensors can be very useful in many SHM applications; however, its advantages over discrete sensors in the context of Bayesian system identification were not as clear initially, since some correlation may be present in the model prediction error of closely space points. Multiple parameter estimations were performed for Bridge 705 using different subsets of discrete and optic fibre sensors, and among all cases, the one that yielded the best match between measurements and posterior model predictions was the one that used only optic fibre data (Section 6.4), proving the potential of this sensing technology for Bayesian inference. In particular, the R^2 scores improved around 8%, 2% and 3% for translations, discrete strains and optic fibre strains, respectively, when compared to a priori model, which already offered reasonably accurate predictions. Still, in order to generalize this conclusion, a more systematic approach that uses a measurement campaign designed for it should be used. Nevertheless, one aspect that should not be overlooked is the qualitative usefulness of quasi-continuous measurements for discovering tendencies, finding erroneous data and making comparisons with model predictions, by building continuous strain plots. This aspect facilitated the application of corrections to the optic fibre measurements described in Appendix B.

C. *How to describe the dependency in model uncertainty for optic fibre sensors and what is their influence on the results of Bayesian parameter estimations?*

The physical model prediction uncertainty was modelled with a multiplicative approach and a squared exponential covariance function. With this data structure, synthetic measurements were generated for Case Study 1, and the ground truth physical parameters used to create them were effectively identified for a 2D case (Section 4.2) and a 5D case (Section 4.4). For Case Study 2, the statistical parameters governing the model uncertainty were not known a priori, so their values were assumed based on previous research on this bridge. Nevertheless, good results were obtained when using these settings for the optic measurements (Section 6.4), as the parameter estimation allowed an improvement of model predictions across all sensor groups. On the other hand, as shown in Section 6.5, inferring the uncertainty parameters along with the physical parameters was unsuccessful, possibly due to the difficulty of obtaining a Gaussian process surrogate that is accurate for log likelihoods with different orders of magnitude, as one obtains when exploring the statistical parameter domain.

Having answered the previous sub-questions, the main research question of this thesis is addressed:

How to perform Bayesian system identification for real-world civil engineering structures within an acceptable running time, i.e., in less than 24 hours, while including high-resolution optic fibre measurements?

An approach that approximates the log joint probability with an actively built Gaussian process was applied to perform multiple Bayesian parameter estimations for both a frame-type structure with synthetic measurements and a cast-in-situ concrete bridge with real measurements. For both cases, the proposed methodology was significantly more efficient on the number of required finite element model evaluations than a direct MCMC sampling, while obtaining equivalent posterior accuracies. Furthermore, cloud-based parallelization was employed to evaluate multiple finite element models simultaneously, decreasing the total running time of the workflow. The combination of these factors made the Bayesian parameter estimation of real-world civil engineering structures with optic fibre measurements feasible for practical situations.

7.2 Contributions

This section summarizes the main contributions of this work, harmonising the research gap found in the literature review and the results obtained for the synthetic and real-world case studies.

First, even though the SHM literature shows some examples of Bayesian system identification using distributed optic fibre measurements, their use is limited to a handful of applications that are either experimental or framed in situations that do not require multiple evaluations of an expensive physical model. To the author's knowledge, this is the first time the combination has been applied for a real civil engineering structure with a computationally demanding finite element model that requires surrogating.

Furthermore, to make the procedure computationally feasible, the log joint probability is replaced with a Gaussian process surrogate using an active posterior learning algorithm. This method requires only a limited number of finite element model evaluations, thus, addressing the main computational bottleneck of the workflow. Another advantage is that the log likelihood, whose calculation scales cubically with the number of data points, no longer needs to be directly obtained since it is approximated by a cheap-to-evaluate Gaussian process surrogate. This technique has been followed in some previous works, although the applications are mostly concentrated in fields such as astrophysics or neuroscience, and the few available civil engineering examples are either theoretical or laboratory-based. To the author's knowledge, this is the first time in open literature where this approach has been used for a real-world civil engineering structure.

Additionally, cloud-based parallelization was successfully used, at a relatively low monetary cost, to perform multiple finite element analyses for both the initial and the active-learning phase of the workflow. By allowing a higher number of model evaluations, this technology has the potential to enhance the use of probabilistic methods in real-world civil structures and increase the dimensionality of parameter estimation problems.

7.3 Future work

The results presented along this document showed the effectiveness of the proposed approach to solve problems that combine Bayesian system identification with optic fibre measurements, giving motivation to continue researching about the topic. This section presents additional research questions for potential upgrades to the methodology, as well as aspects that were not addressed in this thesis.

7.3.1 Improvements on the posterior active learning methodology

- How to build an accurate Gaussian process surrogate of the log joint probability when it drastically varies across the parameter domain, as it happens when inferring statistical parameters?

- How can automatic differentiation be applied to Gaussian processes for obtaining statistical point estimates faster and in a more stable way than with MCMC, in order to evaluate the convergence of the posterior learning procedure?

- What is the effect of parallelization, as introduced in this thesis, on the convergence rate of the posterior learning algorithm? How more advantageous it is to select multiple points per step instead of one?

- How to appropriately balance the degree of exploration and exploitation of the parameter domain when selecting new parallel points for the posterior learning procedure?

- What is the degree of accuracy obtained for posterior predictives when a Gaussian process of the future observations likelihood, which has been fitted only with the samples drawn for the posterior active learning, is used?

7.3.2 Further testing on the use of optic fibre sensors for Bayesian inference

- What is the effect of the spatial discretisation rate of distributed optic fibre strain measurements in the context of Bayesian parameter estimation?

- What is the effect of smoothening/filtering optic fibre measurements on Bayesian parameter estimation and what is the best way to do it in order to maintain theoretical consistency?

- How can dynamic optic fibre strain measurements be incorporated into the Bayesian parameter estimation framework?

- How to build a surrogate of the finite element model strain predictions using the location of the optic fibre data points as a parameter?

References

Acerbi, L., 2018. *Variational Bayesian Monte Carlo.* Montreal, 32nd Conference on Neural Information Processing Systems.

Amsterdam op de kaart, 2022. *Brug 705.* [Online] Available at: https://amsterdamopdekaart.nl/wederopbouw/Cornelis_Lelylaan/Brug_705 [Accessed 06 August 2022].

Ang, A. & Tang, W., 2007. *Probability concepts in engineering: Emphasis on applications to civil and enviromental engineering.* 2nd ed. s.l.:John Wiley and sons Inc..

Bado, M. & Casas, J., 2021. A review of recent distributed optical fiber sensors applications for civil engineering Structural Health Monitoring. *Sensors.*

Bado, M. et al., 2022. Digital twin for civil engineering systems: An exploratory review for distributed sensing updating. *Sensors.*

Bárrias, A., Casas, J. & Villalba, S., 2016. A review of distributed optical fiber sensors for civil engineering applications. *Sensors (Switzerland),* 16(5).

Chai, X., 2019. *Probabilistic system identification and reliability updating for hydraulic structures - Application to sheet pile walls,* Delft: Delft University of Technology.

Chen, H.-P., 2018. *Structural Health Monitoring of Large Civil Engineering Structures.* Hoboken: John Wiley & Sons.

Commenges, D., 2015. Information theory and statistics: an overview. *arXiv preprint arXiv:1511.00860.*

Dashti, M. & Stuart, A. M., 2015. The Bayesian Approach to Inverse-Problems. *ArXiv e-prints.*

Dassault Systemes, 2022. *ABAQUS.* Providence: Software.

DIANA FEA B.V., 2022. *DIANA.* Delft: Software.

Duvenaud, D., 2014. *Automatic model construction with Gaussian processes - PhD dissertation,* Cambridge: University of Cambridge.

Echard, B., Gayton, N. & Lemaire, M., 2011. AK-MCS: An active learning reliability method combining Kriging and Monte Carlo Simulation. *Structural Safety,* Volume 33, pp. 145-154.

Febrianto, E., Butler, L., Girolami, M. & Cirak, F., 2021. Digital twinning of self-sensing structures using the statistical finite-element method. *ArXiv e-prints.*

Fleming, D. P. & VanderPlas, J., 2018. approxposterior: Approximate Posterior Distributions in Python. *The Journal of Open Source Software,* 3(29).

Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J., 2013. emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific,* pp. 306-312.

Gelman, A. et al., 2013. *Bayesian Data Analysis.* 3rd ed. s.l.:CRC Press.

Girolami, M., Febrianto, E., Yin, G. & Cirak, F., 2021. The statistical finite element method (statFEM) for coherent synthesis of observation data and model predictions. *Computer methods in applied mechanics and engineering,* Volume 375.

Goodman, J. & Weare, J., 2010. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science,* 5(1).

Goodman, N. D., Ullman, T. D. & Tenenbaum, J. B., 2011. Learning a Theory of Causality. *Psychological revies,* 118(1).

Harger, F., Garrecht, H. & Lehmann, F., 2019. Structural health monitoring of concrete structures using fibre-optic-based sensors: a review. *Magazine of Concrete Research.*

Hershey, J. & Olsen, P., 2007. *Approximating the Kullback Leibler divergence between Gaussian mixture models.* Honolulu, 2007 IEEE International Conference on Acoustics, Speech and Signal Processing.

Huang, Y. et al., 2019. State-of-the-art review on Bayesian inference in structural system identification and damage assessment. *Advances in Structurel Engineering,* 22(6), pp. 1329-1351.

Huan, X. & Marzouk, Y., 2013. Simulation-based optimal Bayesian experimental design for nonlinear systems. *Journal of Computational Physics,* 232(1), pp. 288-317.

Ierimonti, L. et al., 2021. A transfer Bayesian learning methodology for structural health monitoring of monumental structures. *Engineering Structures,* Volume 247.

Kandasamy, K., Schneider, J. & & Poczos, B., 2015. Bayesian active learning for posterior estimation. *Proceedings of the 24th International Conference on Artificial Intelligence,* pp. 3605-3611.

Katafygiotis, L. & Beck, J., 1998. Updating models and their uncertainties. II: Model identifiability. *Journal of Engineering Mechanics,* 124(4), pp. 463-467.

Kerleguer, B., Cannamela, C. & Garnier, J., 2022. *A Bayesian neural network approach to Multi-fidelity surrogate modelling,* Paris: hal-03608580.

Kish, L., 1965. *Survey sampling.* s.l.:Wiley classics library.

Koune, I., 2021. *Bayesian system identification for structures considering spatial and temporal dependencies - MSc thesis dissertation,* Delft: Delft University of Technology.

Le Gratiet, L., 2013. *Multi-fidelity Gaussian process regression for computer experiments,* Paris: Université Paris-Diderot.

Liu, H., Cai, J. & Ong, Y.-S., 2018. Remarks on multi-output Gaussian process regression. *Knowledge-Based Systems,* Issue 144, pp. 102-121.

McElreath, R., 2016. *Statistical rethinking: a Bayesian course with examples in R and Stan.* Boca Raton: CRC Press/Taylor & Francis Group.

Micchelli, C. A., Xu, Y. & Zhang, H., 2006. Universal kernels. *Journal of Machine Learning Research,* pp. 2651-2667.

Ni, P., Han, Q., Du, X. & Cheng, X., 2022. Bayesian model updating of civil structures with likelihood-free inference approach and response reconstruction technique. *Mechanical Systems and Signal Processing,* Volume 164.

Ni, P. et al., 2021. Probabilistic model updating via variational Bayesian inference and adaptive Gaussian process modeling. *Computational methods in applied mechanics and engineerinig,* Volume 383.

Osborne, M. et al., 2012. Active learning of model evidence using Bayesian quadrature. *Advances in neural information processing systems,* pp. 46-54.

Pepi, C., Gioffre, M. & Grigoriu, M., 2019. Parameters identification of cable-stayed footbridges using Bayesian inference. *Meccanica,* Volume 54, pp. 1403-1419.

Rasmussen, C. & Williams, C., 2006. *Gaussian processes for machine learning.* s.l.:MIT Press.

Rózsás, Á., Slobbe, A., Martini, G. & Jansen, R., 2022. Structural and load parameter estimation of a real-world reinforced concrete slab bridge using measurements and Bayesian statistics. *Structural Concrete,* pp. 1-32.

Sbarufatti, C., Cadini, F., Locatelli, A. & Giglio, M., 2018. *Surrogate modelling for observation likelihood calculation in a particle filter framework for automated diagnosis and prognosis.* Manchester, 9th European Workshop on Structural Health Monitoring.

Schranz, B. W. P., Czaderski, C. & Shahverdi, M., 2022. Fibre optic measurements and model uncertainty quantification for Fe-SMA strengthened concrete structures. *Engineering Structures,* Volume 256.

Simoen, E., C., P., De Roeck, G. & G., L., 2012. *Influence of the prediction error correlation model on Bayesian FE model updating results.* s.l., Life-Cycle and Sustainability of Civil Infrastructure Systems - Proceedings of the 3rd International Symposium on Life-Cycle Civil Engineering, IALCCE 2012.

Simoen, E., Papadimitriou, C. & Lombaert, G., 2013. On prediction error correlation in Bayesian model updating. *Journal of Sound and Vibration,* 332(18), pp. 4136-4152.

Skilling, J., 2006. Nested sampling for general Bayesian computation. *Bayesian Analysis,* 1(4), pp. 833-860.

Speagle, J. S., 2020. *A Conceptual Introduction to Markov Chain Monte Carlo Methods,* Cambridge: Harvard & Smithsonian Center for Astrophysics.

Speagle, J. S., 2020. DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Monthly Notices of the Royal Astronomical Society,* pp. 3132-3158.

TNO, 2014. *Extending the Service Life of Civil Structures,* s.l.: Netherlands Organisation for Applied Scientific Research.

TNO, 2022. *Taralli: a parameter estimation tool for statistical problems,* s.l.: Netherlands Organisation for Applied Scientific Research.

Torzoni, M., Manzoni, A. & Mariani, S., 2022. Health monitoring of civil structures: A MCMC approach based on multi-fidelity deep neural network surrogate. *Computer sciences and mathematics forum,* 16(2).

Waeytens, J. et al., 2016. Model updating techniques for damage detection in concrete beam using optical fiber strain measurement device. *Engineering Structures,* Volume 129, pp. 2-10.

Wang, H. & Li, J., 2017. Adaptive Gaussian process approximation for Bayesian inference with expensive likelihood functions. *ArXiv e-prints.*

Wang, H. et al., 2016. Surrogate-assisted Bayesian inference inverse material identification method and application to advanced high strength steel. *Inverse Problems in Science and Engineering,* 24(7), pp. 1133-1161.

Wosniok, A. et al., 2019. *Static load monitoring of a concrete bridge using a high-precision distributed fiber optic sensor system.* Postdam, SMAR 2019 - Fifth Conference on Smart Monitoring Assessment and Rehabilitation of Civil Structures.

Wu, L. et al., 2020. Bayesian inference of non-linear multiscale model parameters accelerated by a Deep Neural Network. *Computational methods in applied mechanics and engineering,* Volume 360.

Wu, T., Liu, G., Fu, S. & Xing, F., 2020. Recent progress of fiber-optic sensors for the Structural Health Monitoring of civil infrastructure. *Sensors.*

Zhang, Y. & Yang, W., 2013. Bayesian strain modal analysis under ambient vibration and damage identification using distributed fiber Bragg grating sensors. *Sensors and Actuators A: Physical,* Volume 201, pp. 434-449.

Zonta, D. et al., 2013. Sensor Fusion on Structural Monitoring Data Analysis: Application to a Cable-Stayed Bridge. *Key Engineering Materials,* Volume 569, pp. 812-819.

Appendix A – Implementation details

Python 3.8 was used to perform the computations of this thesis. Table 35 and Table 36 show both the Python and external tools employed in Case studies 1 and 2, respectively.

Table 35: Tools used for Case Study 1 – Synthetic frame.

Tool	Description	Use
NumPy	Python library that provides fast operations for arrays, including mathematical, logical, sorting, etc.	General computing
SciPy	Python library that provides fundamental algorithms for scientific computing	General computing
Matplotlib	Python library for creating static, animated and interactive visualizations	Plotting
emcee	Python package that implements the Affine Invariant Markov chain Monte Carlo (MCMC) Ensemble sampler	MCMC sampling
Taralli	Python package that provides functionalities for Bayesian system identification	Compiling covariance matrices
approxposterior	Python package that implements a variant of the Bayesian Active Learning for Posterior estimation algorithm	Building a GP surrogate of the joint probability

Table 36: Tools used for Case Study 2 – Bridge 705.

Tool	Description	Use
NumPy	Python library that provides fast operations for arrays, including mathematical, logical, sorting, etc.	General computing
SciPy	Python library that provides fundamental algorithms for scientific computing	General computing
Matplotlib	Python library for creating static, animated and interactive visualizations	Plotting
emcee	Python package that implements the Affine Invariant Markov chain Monte Carlo (MCMC) Ensemble sampler	MCMC sampling
Taralli	Python package that provides functionalities for Bayesian system identification	Compiling covariance matrices
scikit-learn	Python library for machine learning featuring various regression, classification and clustering algorithms	Building a GP surrogate of the joint probability
Diana FEA	Software for Finite Element Analysis of civil engineering structures	Finite element analysis
Microsoft Azure	Cloud computing service for application management via Microsoft-managed data centres	Cloud-based parallelization of FE analyses
Kubernetes	Open-source container orchestration system for automating software deployment, scaling and management	Cloud-based parallelization of FE analyses
Argo Workflows	Open-source container-native workflow engine for orchestrating parallel jobs on Kubernetes	Cloud-based parallelization of FE analyses

Appendix B – Optic fibre corrections

In this section, the corrections that were applied to the optic fibre measurements for each load case are presented. Notice that S12-ZB is not displayed since there are no available recordings for this load case.

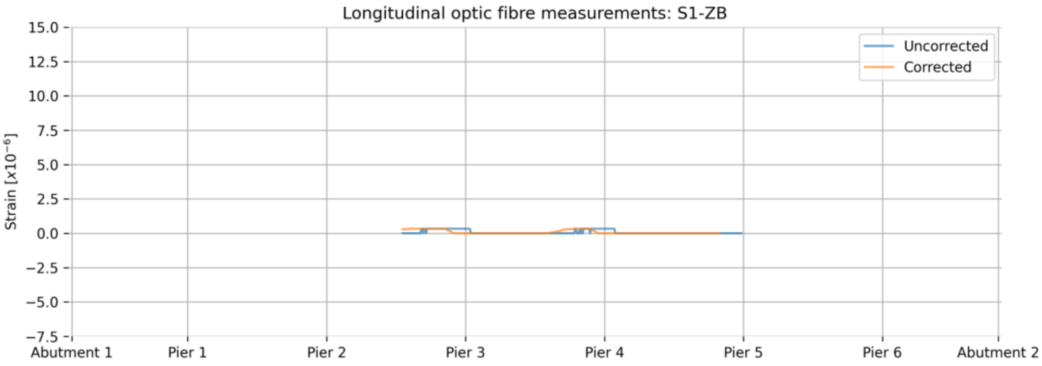
When comparing the uncorrected measurements to the model predictions, it was evident that the shapes were similar but there was an offset between them. After consulting with the team that conducted the campaign, a 2.5 m shift was applied to the measurements of the bottom leg with respect to the optic fibre coordinates, while a 1.5 m shift was applied to the transversal leg if the optic fibre. The effect of the shifts can clearly be seen for load cases S5-ZB to S9-ZB.

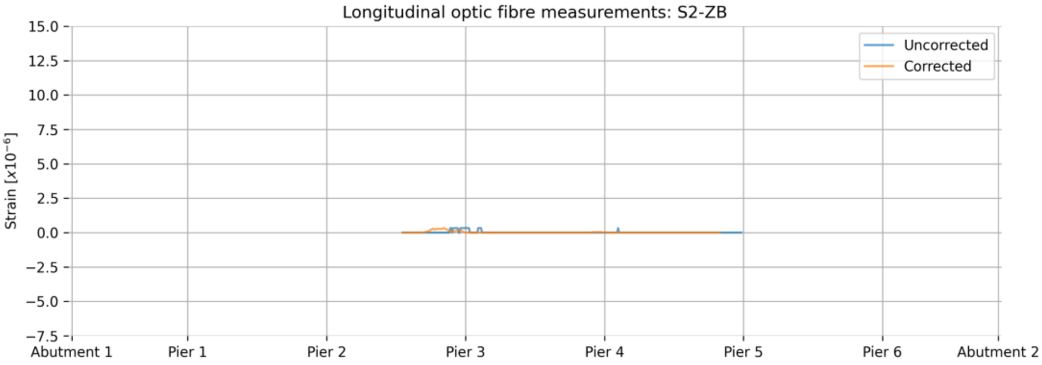
In load cases such as S10-ZB, S11-ZB and S13-ZB, the truck position is so far from the optic fibre that the model predicts strain values close to zero; however the plots show almost constant non-zero values in these areas. It was found that a small temperature change had occurred during the test series. Since temperature loading was not included in the finite element model, the corrections shown Table 37 below were necessary to match the model predictions:

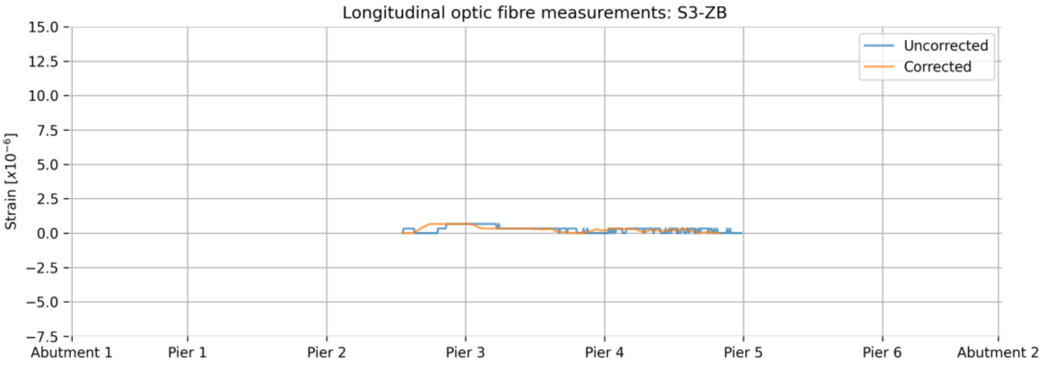
Table 37: Temperature corrections applied to the optic fibre measurements.

Time event	Load case	ΔT (Cº)	Temperature correction
24-Oct-2018 00:53:00	S1-ZB	0.00	0.0 microstrains
24-Oct-2018 01:00:00	S2-ZB	0.00	0.0 microstrains
24-Oct-2018 01:07:00	S3-ZB	0.00	0.0 microstrains
24-Oct-2018 01:14:00	S4-ZB	0.00	0.0 microstrains
24-Oct-2018 01:18:00	S5-ZB	0.00	0.0 microstrains
24-Oct-2018 01:22:00	S6-ZB	0.00	0.0 microstrains
24-Oct-2018 01:26:00	S7-ZB	0.02	2.4 microstrains
24-Oct-2018 01:30:00	S8-ZB	0.03	3.6 microstrains
24-Oct-2018 01:34:00	S9-ZB	0.03	3.6 microstrains
24-Oct-2018 01:38:00	S10-ZB	0.04	4.8 microstrains
24-Oct-2018 01:41:00	S11-ZB	0.04	4.8 microstrains
24-Oct-2018 01:48:00	S13-ZB	0.06	7.2 microstrains

Finally, the raw measurements were replaced by moving averages centred at the measurement location with a window of 10 data points (around 1.0 m). This was necessary since the data exhibited a peakiness that appeared to depend of the intensity of the measurement itself, so it could not be modelled with the data generating process described in Section 3.2. This issue can be observed especially for load cases S5-ZB, S7-ZB and S9-ZB.







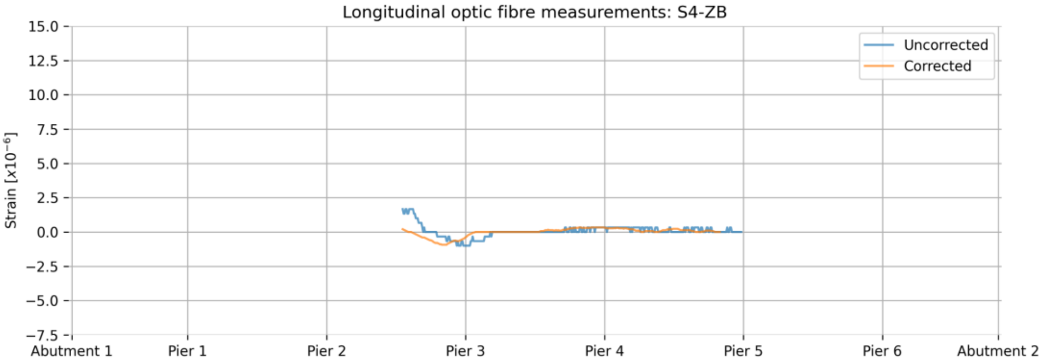


Figure 59: Comparison of corrected and uncorrected optic fibre measurements in the longitudinal direction of the bridge for load cases S1-ZB, S2-ZB, S3-ZB and S4-ZB.

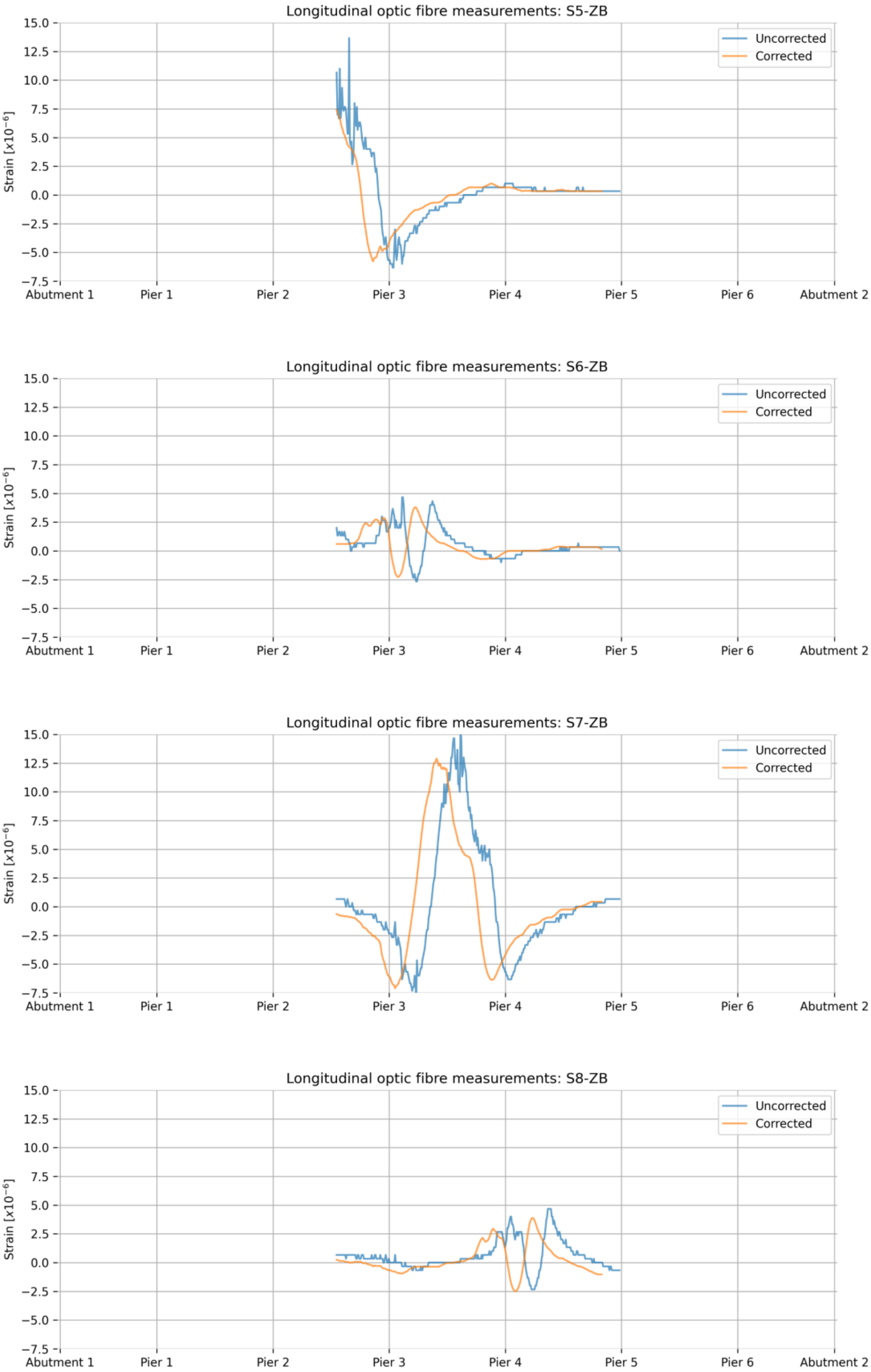


Figure 60: Comparison of corrected and uncorrected optic fibre measurements in the longitudinal direction of the bridge for load cases S5-ZB, S6-ZB, S7-ZB and S8-ZB.

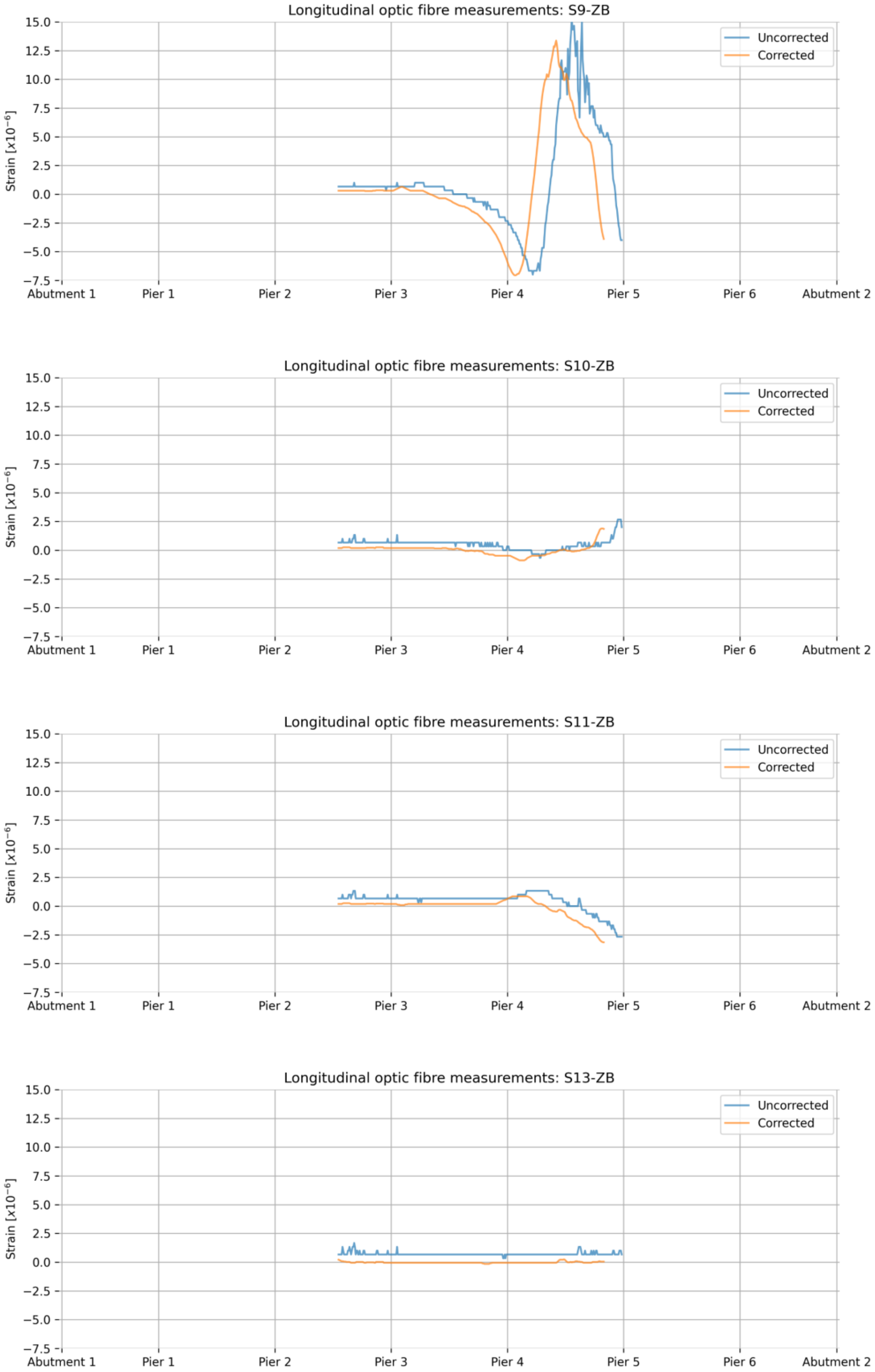


Figure 61: Comparison of corrected and uncorrected optic fibre measurements in the longitudinal direction of the bridge for load cases S9-ZB, S10-ZB, S11-ZB and S13-ZB.

Appendix C – Sub-case 1 additional results

The goal of Sub-case 1 was to validate the methodology proposed in this thesis, so the emphasis of Section 6.2 was put on the posterior distributions. This section presents additional results related to the fitness of the a posteriori model predictions.

The posterior median values from Table 22 are used to obtain finite element predictions, which are then compared to the sensor measurements in Figure 62. A clear improvement on the translation predictions are obtained with respect to the a priori model, as can be observed by an R^2 score increment of around 9% and by the fact that the 0.99 credible intervals cross the diagonal line for every dot.

On the other hand, the predictive capacity is not considerably improved for the discrete strains and it is even worse for sensor group 'T'. However, this subset has been previously identified as suspicious and its measurements may not be entirely reliable (notice that the credible interval do not cross the diagonal for multiple dots).

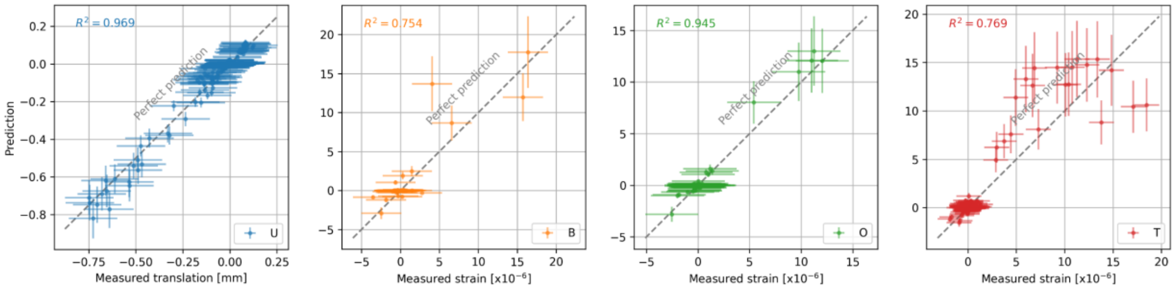


Figure 62: Comparison of model predictions obtained with Sub-case 1 converged posterior median values (y-axis) vs measurements (x-axis) for translation ('U') and strain ('B', 'O' and 'T') sensor groups. The vertical and horizontal lines on the dots indicate the 0.99 highest density credible intervals for model prediction uncertainty and measurement uncertainty, respectively.

Even though the optic fibre measurements were not used for this Sub-case, it is still interesting to notice that the R^2 score is 9% lower than the one from the a priori model, as it can be seen in Figure 63.

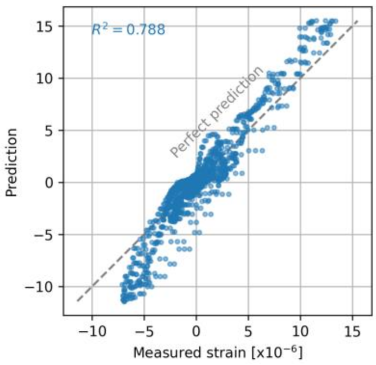


Figure 63: Comparison of model predictions obtained with Sub-case 1 converged posterior median values (y-axis) vs measurements (x-axis) for the optic fibre sensor.

The findings described in the previous paragraphs can also be visualized in the following response diagrams where the model prediction uncertainty is represented as a grey shadowed region while the measurement uncertainty of the optic fibre and discrete measurements are displayed as a light orange shadowed region and vertical black lines, respectively.

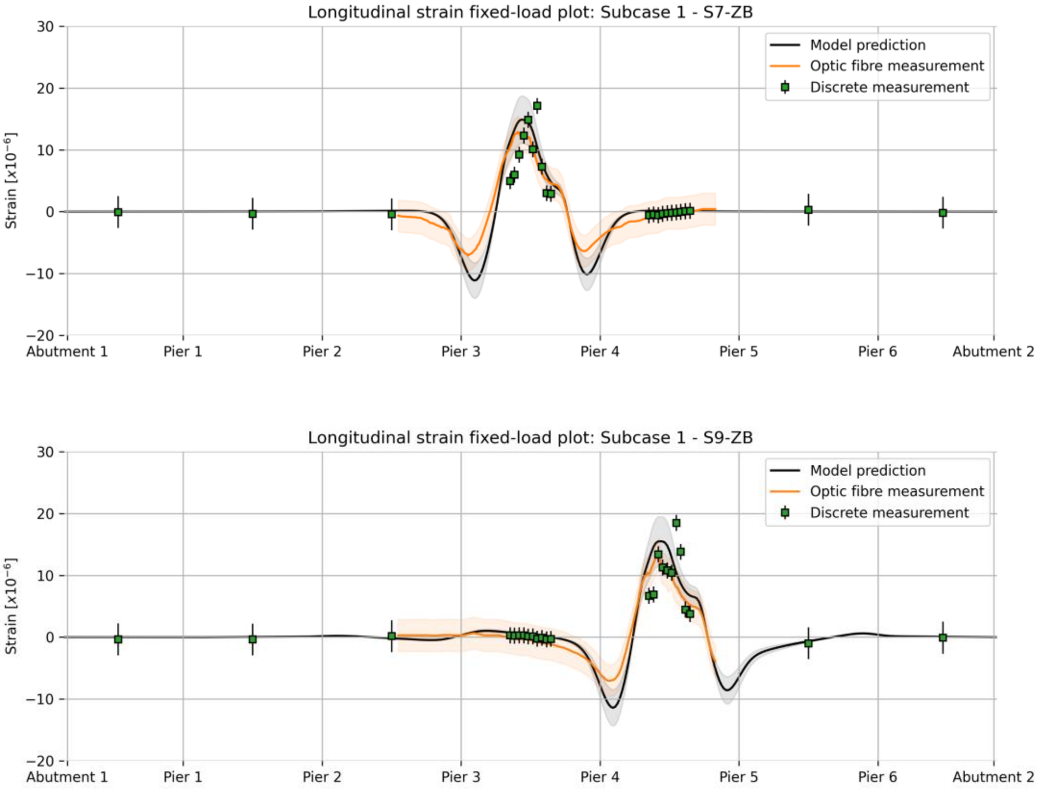
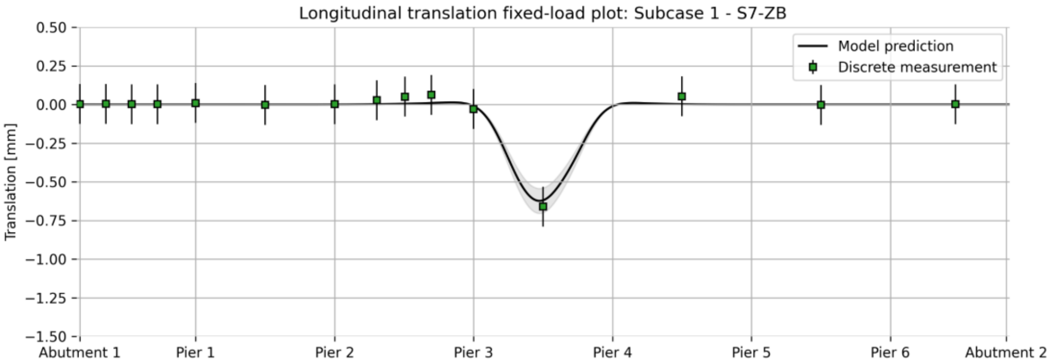


Figure 64: Longitudinal strain diagrams at centreline of Truck 1 lane (deck bottom fibre). The model predictions are obtained with the posterior median parameters from Sub-case 1. The grey-shaded region indicate the 0.99 highest density credible intervals for the model prediction uncertainty. The orange-shaded region and black vertical lines indicate the 0.99 highest density credible intervals for measurement uncertainty for optic fibre and discrete measurements, respectively. Top: load case S7-ZB. Bottom: load case S9-ZB.

The strain diagrams from Figure 64 show that the predictions are farther to the measurements than for the a priori model (Figure 36), especially in the peaks. Also, the 0.99 measurement uncertainty credible interval do not cross the model prediction for many discrete sensors. This confirms the reduction of R^2 scores for the strain sensors. On the other hand, Figure 65 shows a remarkable agreement between the predicted translations and measurements, especially in the midspan peaks, which is also backed by the R^2 score increment for sensor group ‘U’.



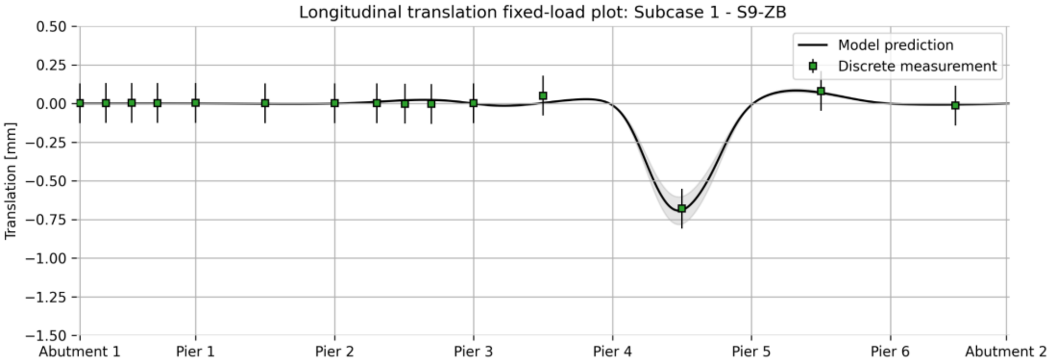


Figure 65: Vertical translation diagrams at centreline of Truck 1 lane. The model predictions are obtained with the posterior median parameters from Sub-case 1. The grey-shaded region indicate the 0.99 highest density credible intervals for the model prediction uncertainty. The orange-shaded region and black vertical lines indicate the 0.99 highest density credible intervals for measurement uncertainty for optic fibre and discrete measurements, respectively. Top: load case S7-ZB. Bottom: load case S9-ZB.

Appendix D – Sub-case 2 additional results

The goal of Sub-case 2 was to assess the a posteriori model predictions obtained using a subset of discrete strain sensors, so the emphasis of Section 6.3 was put on showing the discrepancies between model predictions and measurements via R^2 score plots and response plots. This section presents additional results related to the active sampling process.

The evolution of the parameter estimation per step for Sub-case 2 is presented in Figure 66. It can be seen from the top plots that the median and MADs show a stable flat behaviour only in the last steps. This is confirmed by the bottom sub-figure, where the convergence criterion, i.e., the median variation is below the threshold in three consecutive steps, is achieved in the last step displayed.

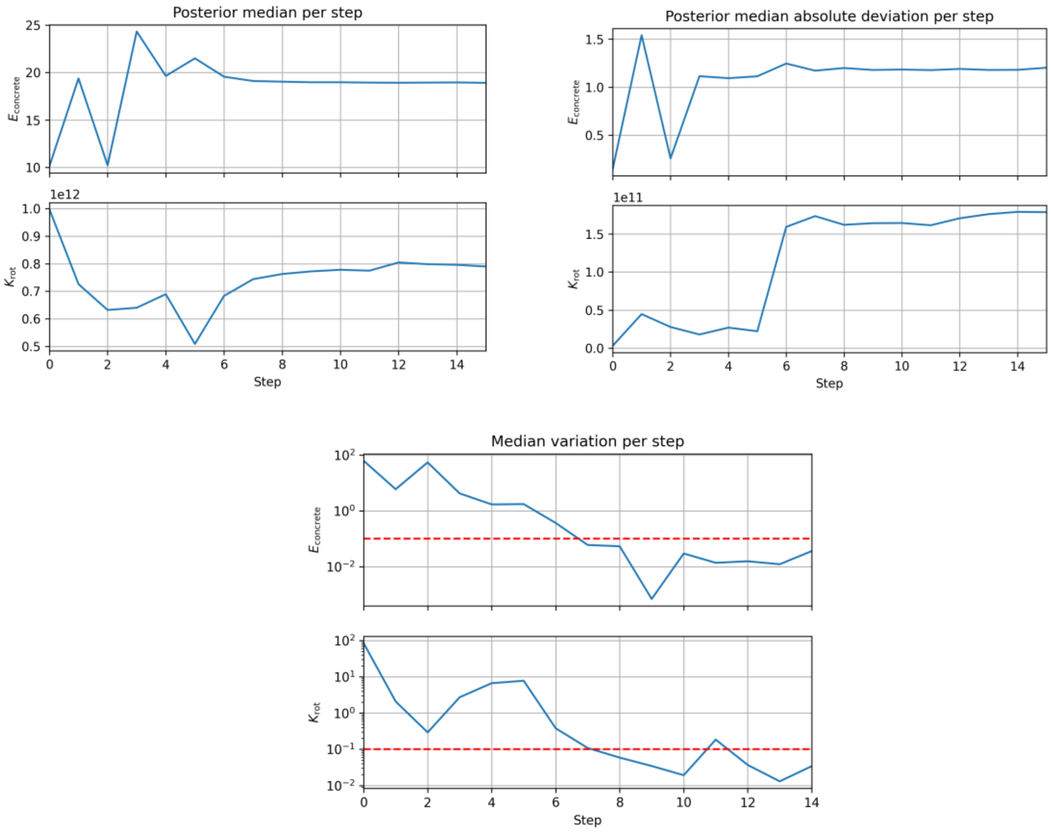


Figure 66: Evolution of MCMC parameter estimation per step for Sub-case 2, with four new samples drawn at each of them. Step 0 only uses the initial samples, i.e., before the active learning. Top left: Median values. Top right: Maximum absolute deviations. Bottom: Variation of the mean with respect to previous step normalized by the MAD. The red dashed line represents the convergence threshold set for this case study: 0.10.

The distribution of actively drawn samples from Figure 67 is shifted to the left of Sub-case 1 and is also less concentrated, mainly due to the fact that that many new points on the lower boundary for E_c and in both for K_{rot} were chosen. As it has been explained before, the GP surrogate tends to overestimate the log joint probabilities on the borders of the domain, so the algorithm initially predicts that this regions will have high probability densities and selects points on it. This is then corrected but convergence is achieved at a total of 80 finite element model evaluations, 24 more than in Sub-case 1.

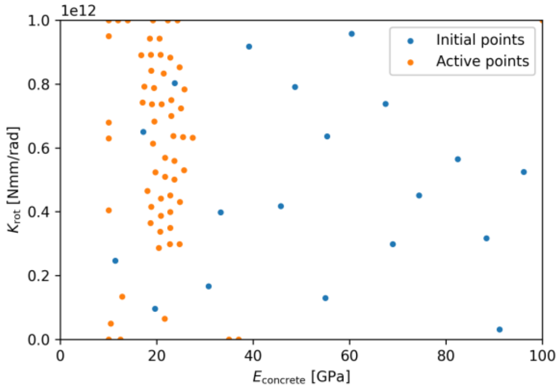


Figure 67: Samples used to surrogate the log joint probability GP surrogate for Sub-case 2. The blue dots are the initial Latin hypercube points while the orange are actively drawn.

Appendix E – Sub-case 3 additional results

The goal of Sub-case 3 was to assess the a posteriori model predictions obtained using the optic fibre sensor, so the emphasis of Section 6.4 was put on showing the discrepancies between model predictions and measurements via R^2 score plots and response plots. This section presents additional results related to the active sampling process.

The evolution of the parameter estimation per step for Sub-case 3 is presented in Figure 68. Convergence is achieved at step 9, which is when three consecutive step have a median variation below the 0.1 threshold for the two parameters simultaneously. The posterior values remain quite flat after this point except for the MAD of E_c, which exhibits some small fluctuations below 0.05 GPa.

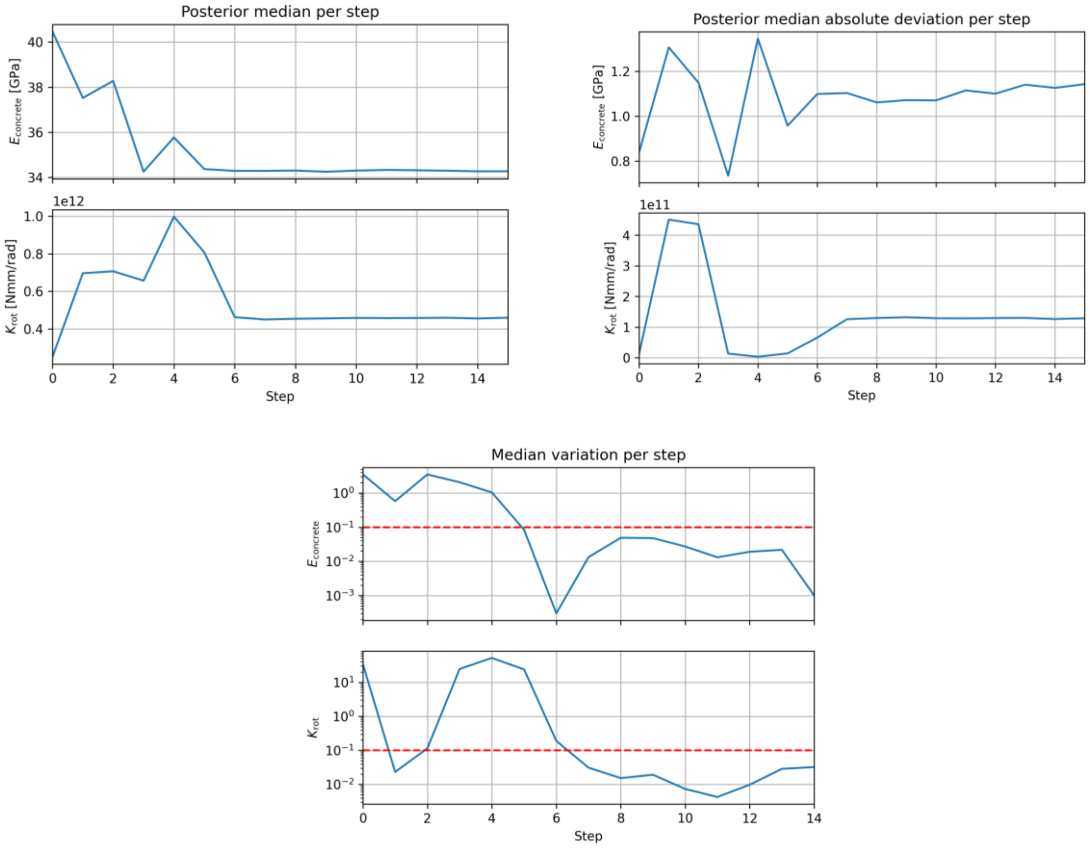


Figure 68: Evolution of MCMC parameter estimation per step for Sub-case 3. Step 0 only uses the initial samples, i.e., before the active learning. Top left: Median values. Top right: Maximum absolute deviations. Bottom: Variation of the mean with respect to previous step normalized by the MAD. The red dashed line represents the convergence threshold set for this case study: 0.10.

As with the previous cases, the actively drawn samples for Sub-case 3 shown in Figure 69 are concentrated in a narrow region of E_c centred around its converged posterior median value, while they are more evenly distributed for K_{rot}. Here again, the algorithm selected many points on the boundaries of the domain that ended up having very low probability densities, due to a GP surrogate overestimated prediction at the borders of the parameter space.

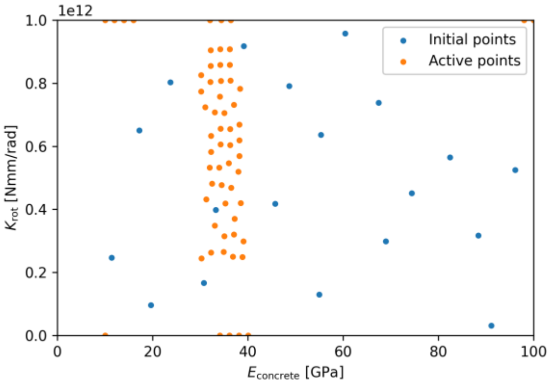


Figure 69: Samples used to surrogate the log joint probability GP surrogate for Sub-case 3. The blue dots are the initial Latin hypercube points while the orange are actively drawn.

Appendix F – A priori fixed-load plots

A complete list of all strain and translation response plots for the a priori model are presented in this section.

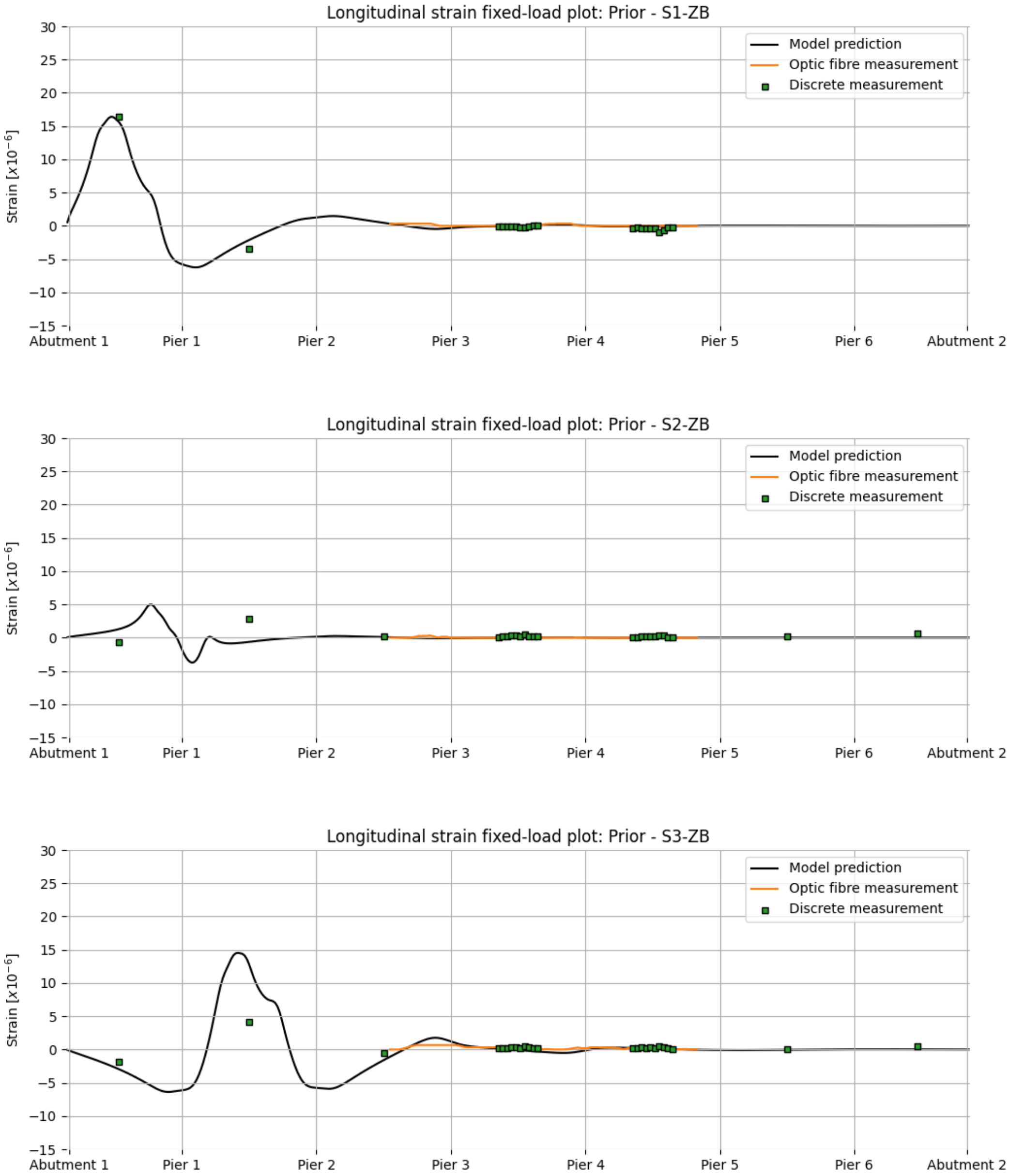
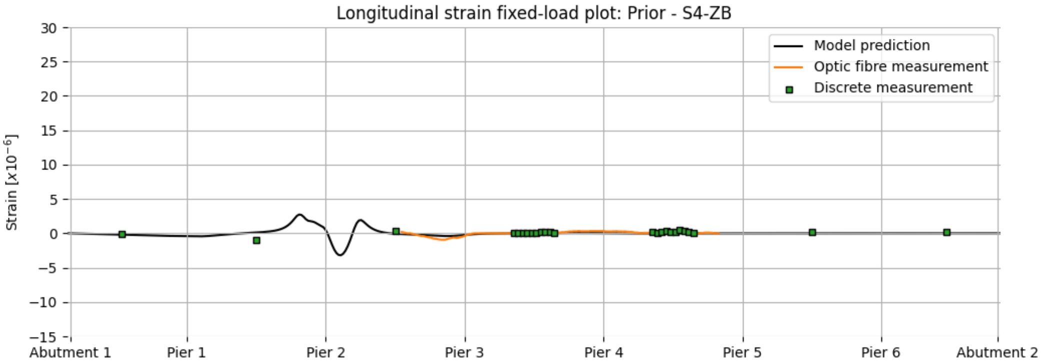
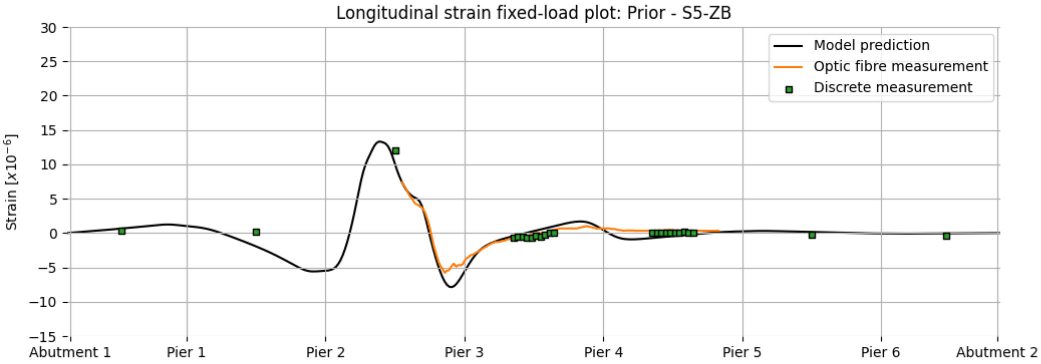
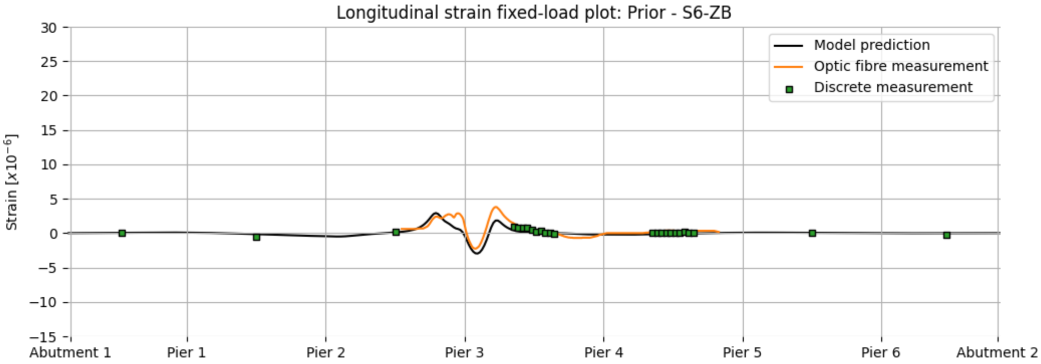


Figure 70: Longitudinal strain diagrams at centreline of Truck 1 lane (deck bottom fibre) for load cases S1-ZB, S2-ZB and S3-ZB. The model predictions are obtained with the a priori model parameters.







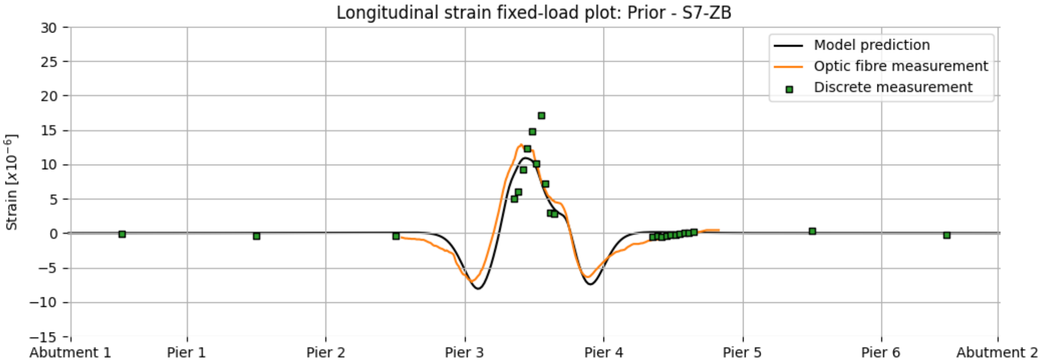


Figure 71: Longitudinal strain diagrams at centreline of Truck 1 lane (deck bottom fibre) for load cases S4-ZB, S5-ZB, S6-ZB and S7-ZB. The model predictions are obtained with the a priori model parameters.

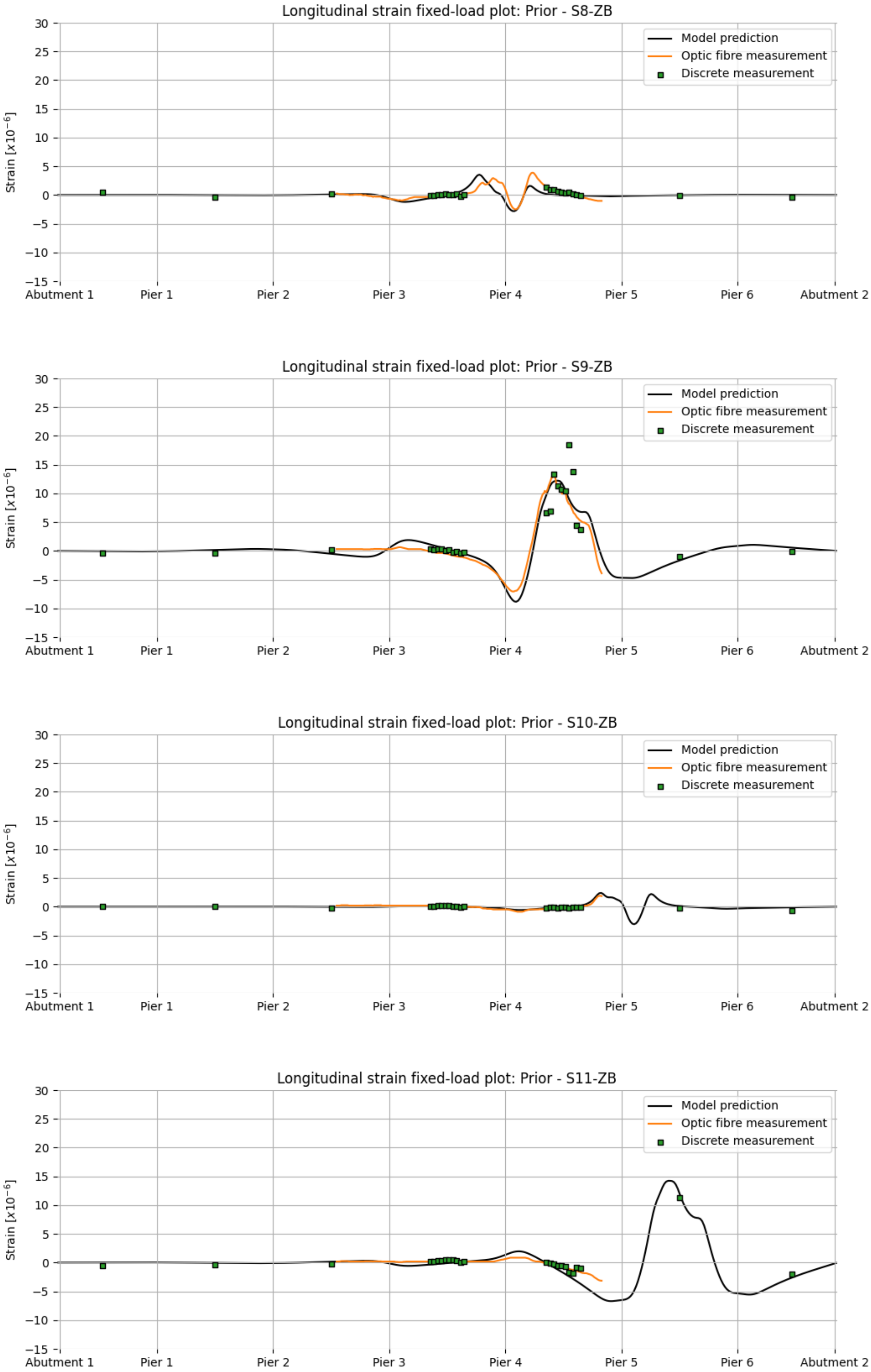
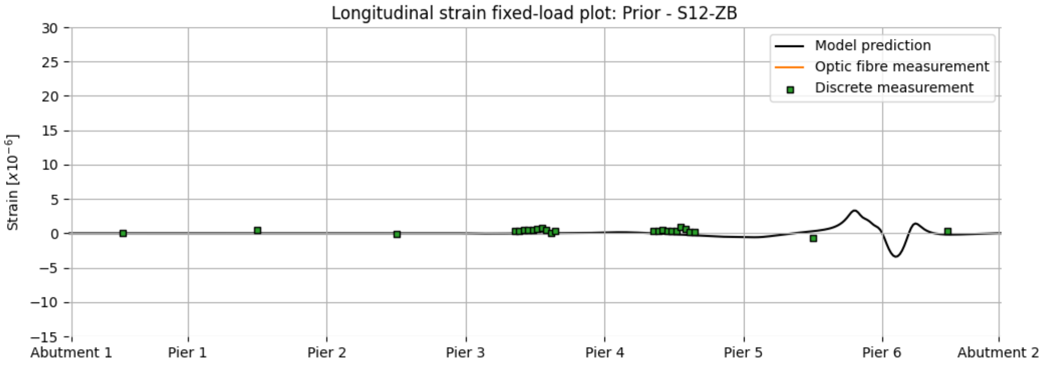


Figure 72: Longitudinal strain diagrams at centreline of Truck 1 lane (deck bottom fibre) for load cases S8-ZB, S9-ZB, S10-ZB and S11-ZB. The model predictions are obtained with the a priori model parameters.



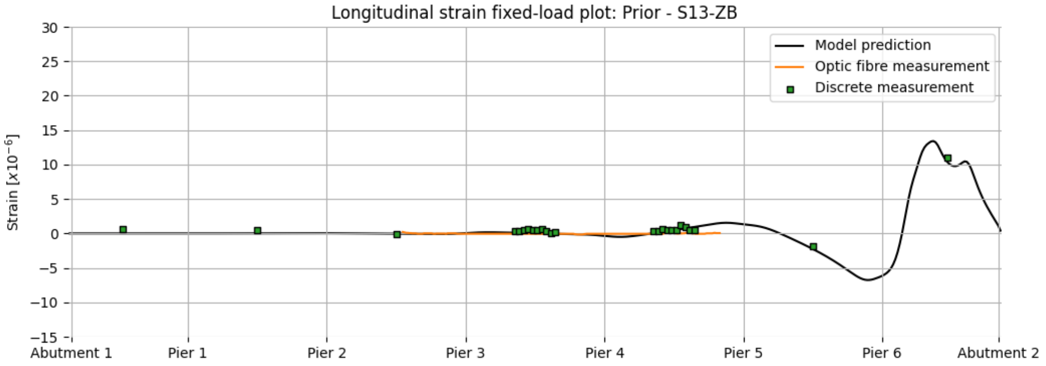
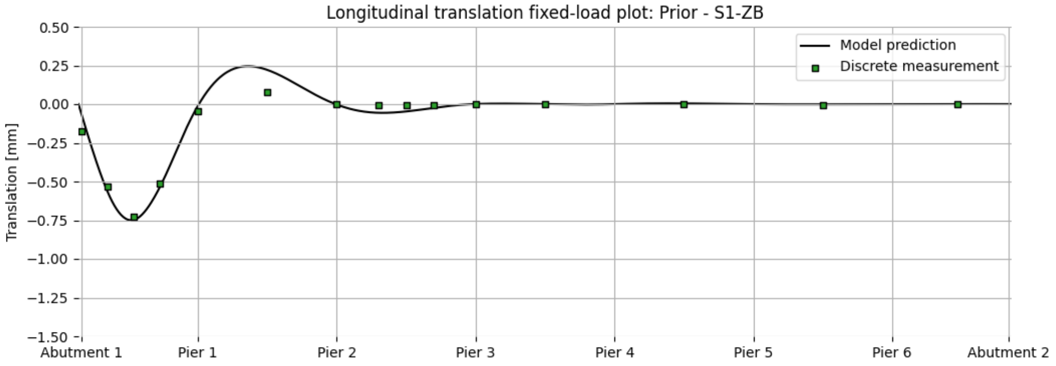


Figure 73: Longitudinal strain diagrams at centreline of Truck 1 lane (deck bottom fibre) for load cases S12-ZB and S13-ZB. The model predictions are obtained with the a priori model parameters.



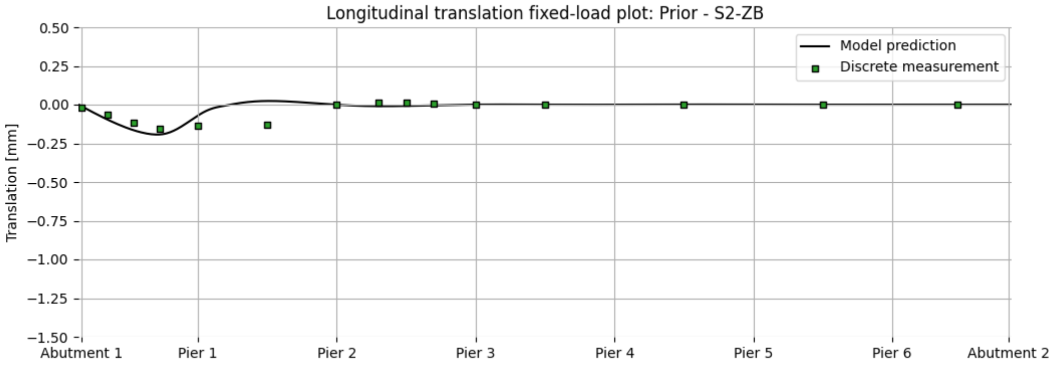
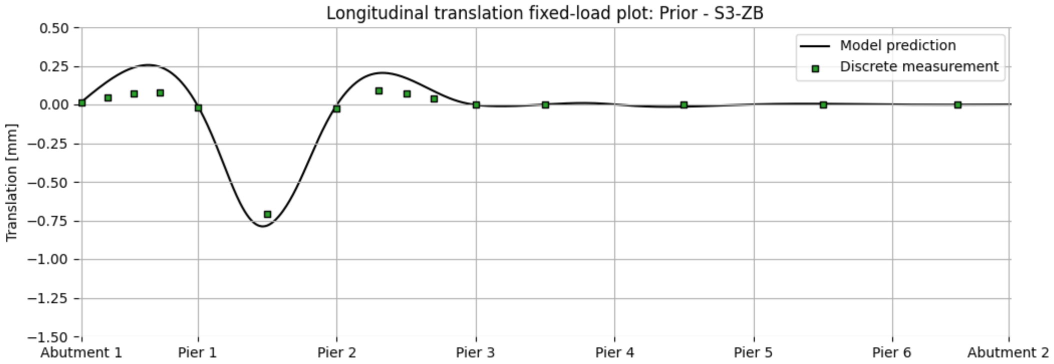
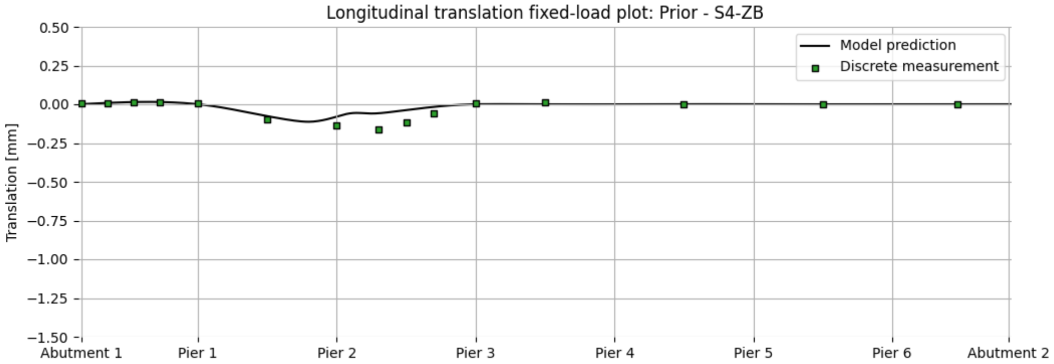
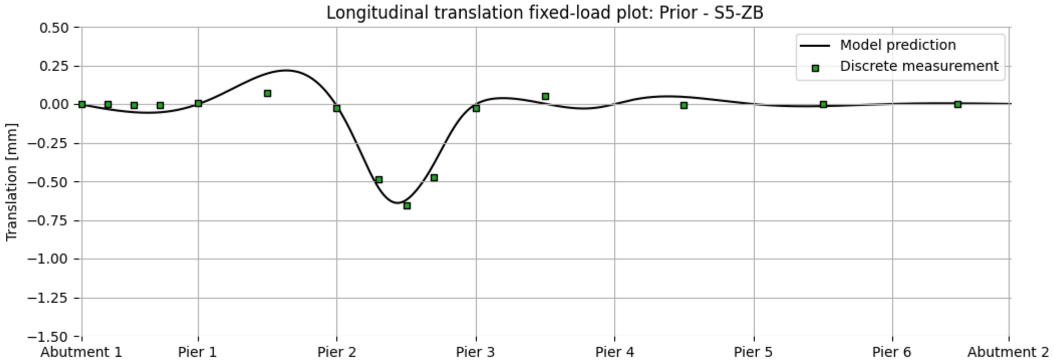


Figure 74: Vertical translation diagrams at centreline of Truck 1 lane for load cases S1-ZB and S2-ZB. The model predictions are obtained with the a priori model parameters.







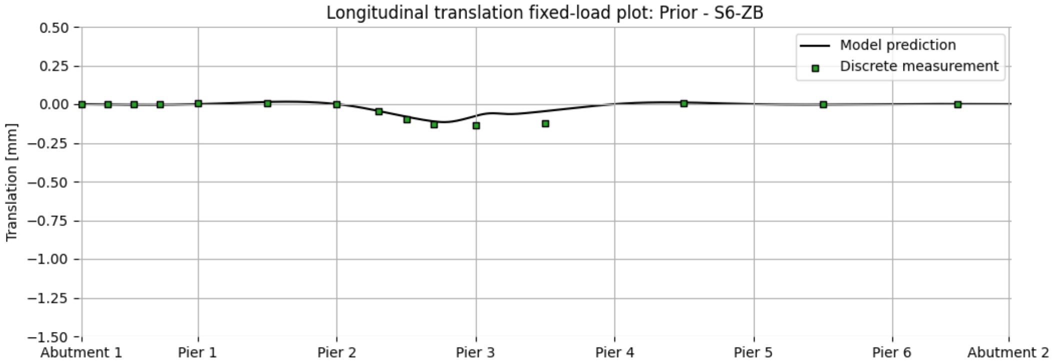


Figure 75: Vertical translation diagrams at centreline of Truck 1 lane for load cases S3-ZB, S4-ZB, S5-ZB and S6-ZB. The model predictions are obtained with the a priori model parameters.

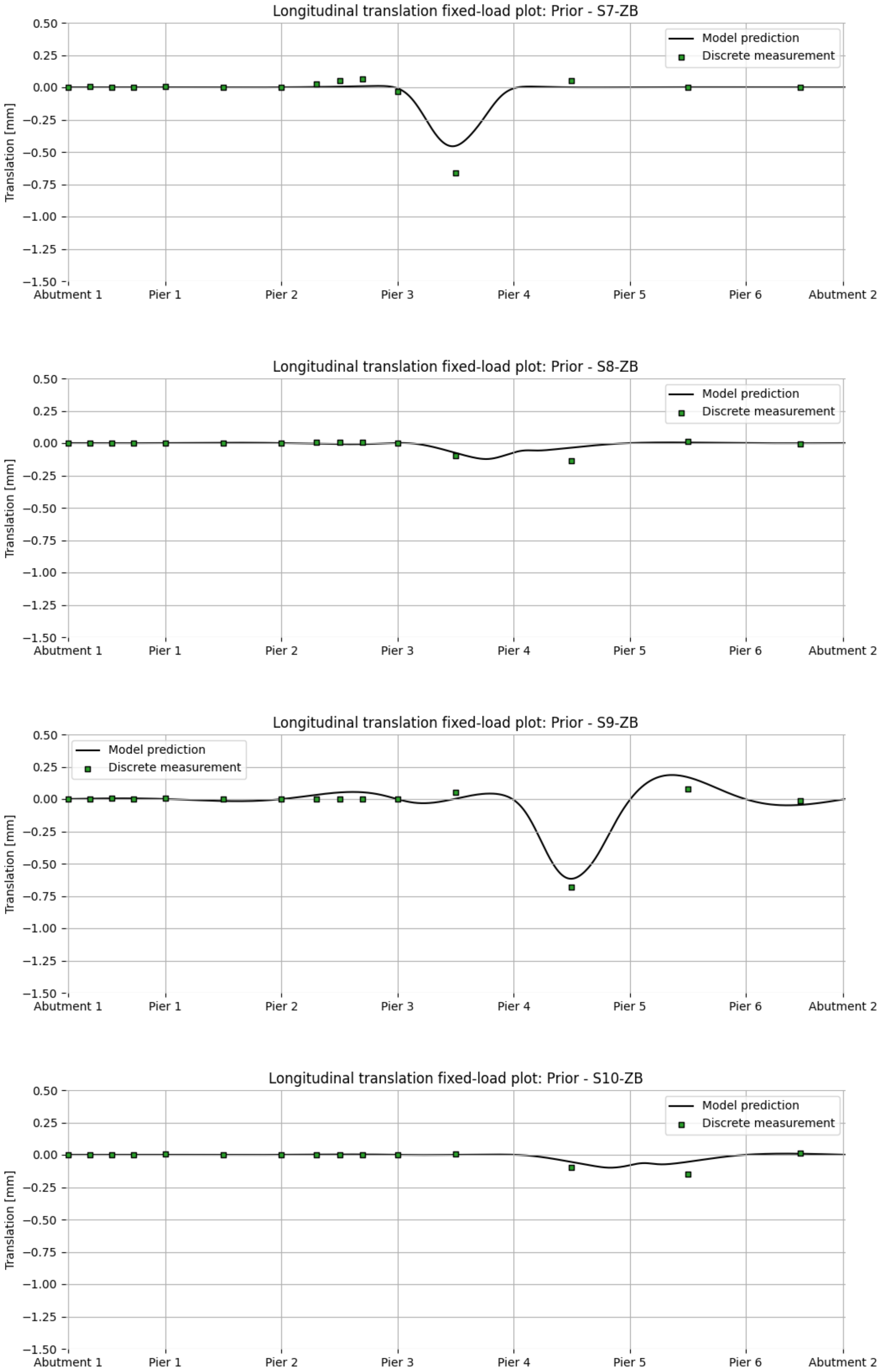


Figure 76: Vertical translation diagrams at centreline of Truck 1 lane for load cases S7-ZB, S8-ZB, S9-ZB and S10-ZB. The model predictions are obtained with the a priori model parameters.

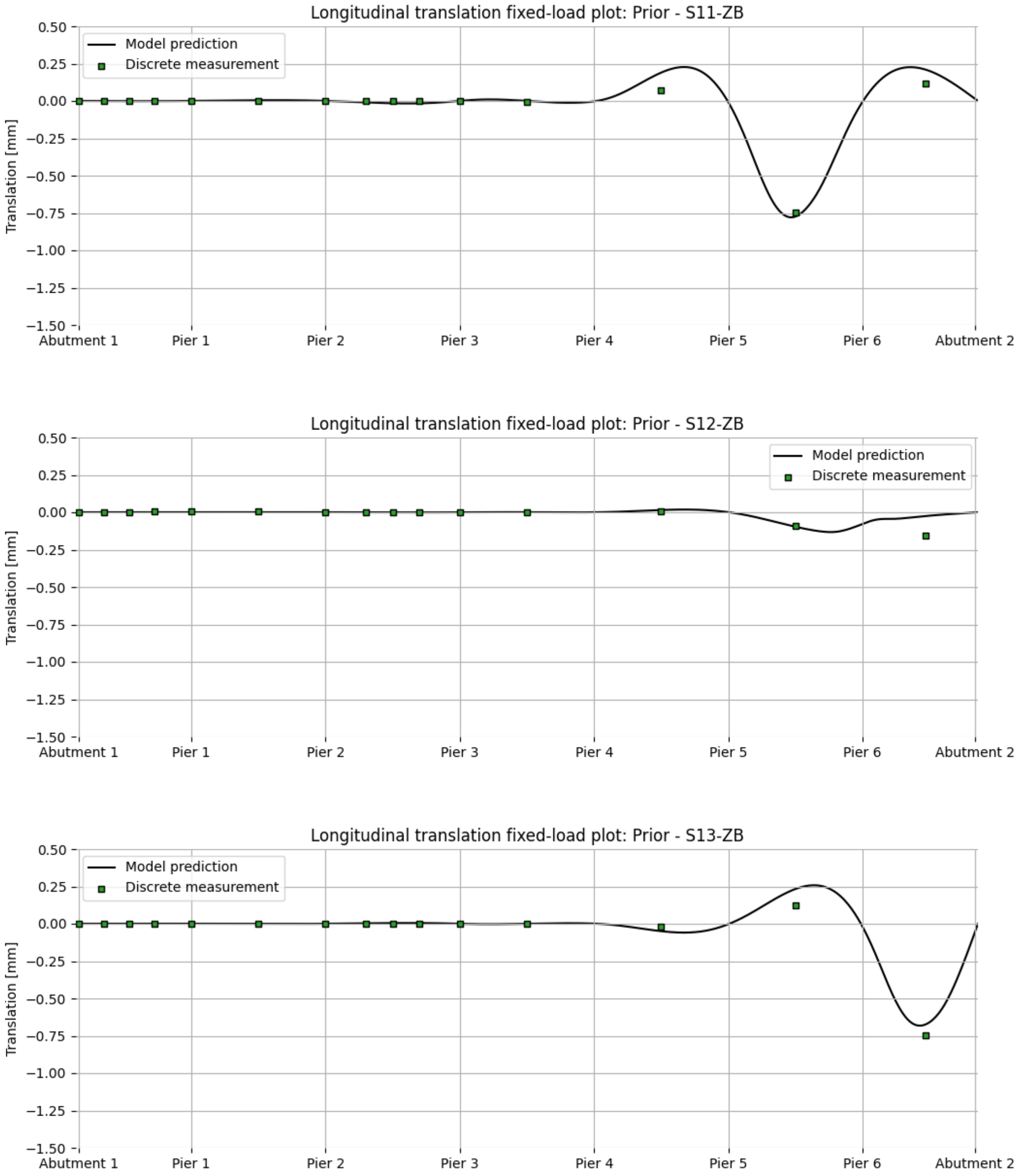


Figure 77: Vertical translation diagrams at centreline of Truck 1 lane for load cases S11-ZB, S12-ZB and S13-ZB. The model predictions are obtained with the a priori model parameters.

Appendix G – Sub-case 3 fixed-load plots

A complete list of all strain and translation response plots for Sub-case 3 are presented in this section. The grey-shadowed region corresponds to the 0.99 highest density credible interval for model predictions, which are obtained by multiplying the response prediction, the strain model uncertainty coefficient of variation (0.10) and the factor associated with a 0.99 credible interval for a normal distribution (2.58). The orange region and vertical black lines correspond to the 0.99 highest density credible interval for measurements of the optic fibre and discrete sensors, respectively.

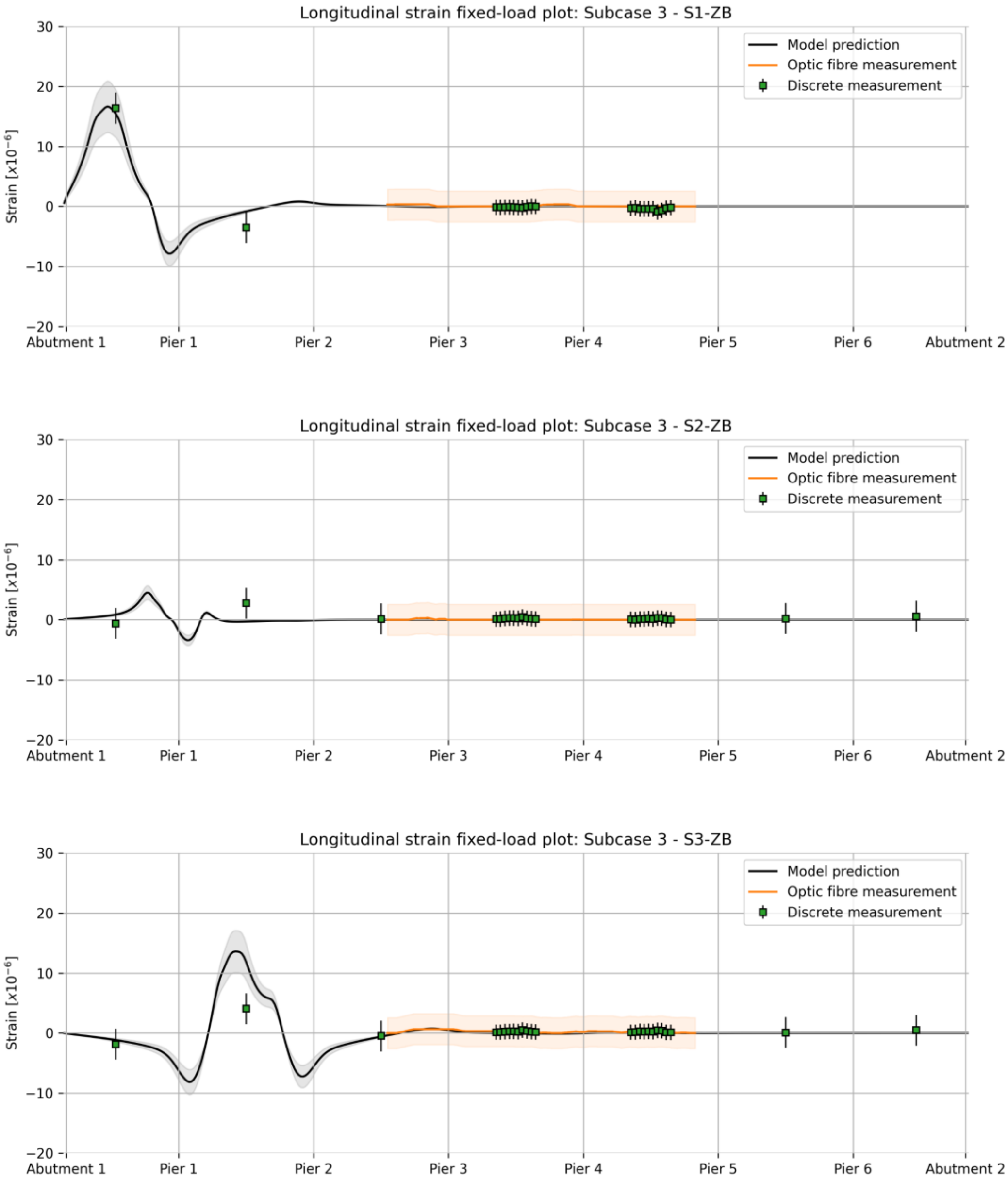


Figure 78: Longitudinal strain diagrams at centreline of Truck 1 lane (deck bottom fibre) for load cases S1-ZB, S2-ZB and S3-ZB. The model predictions are obtained with the posterior median Sub-case 3 parameters.

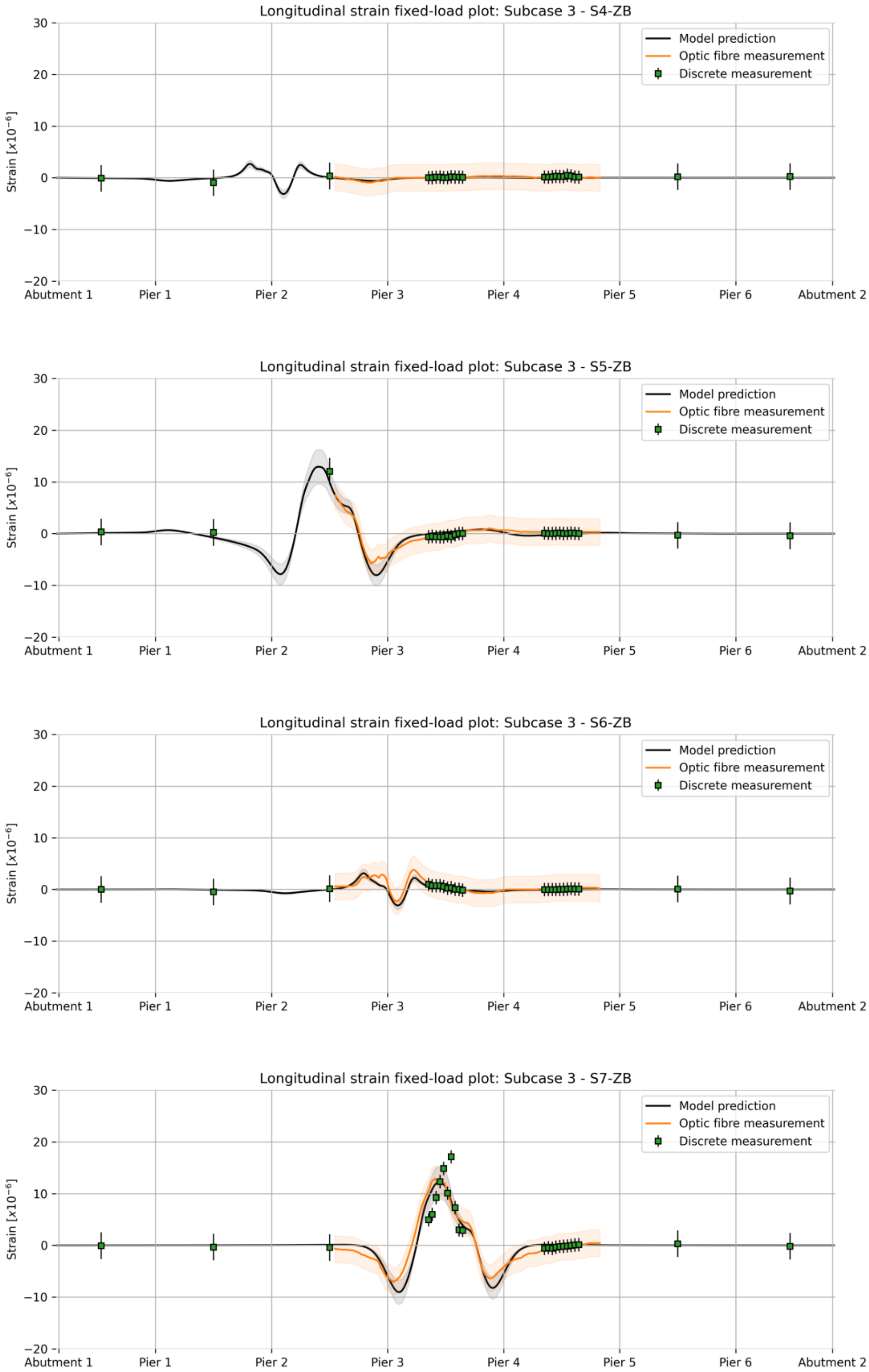
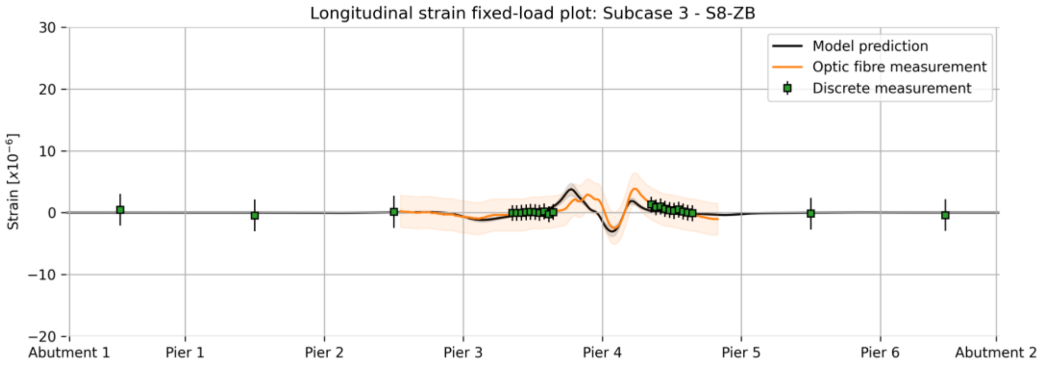
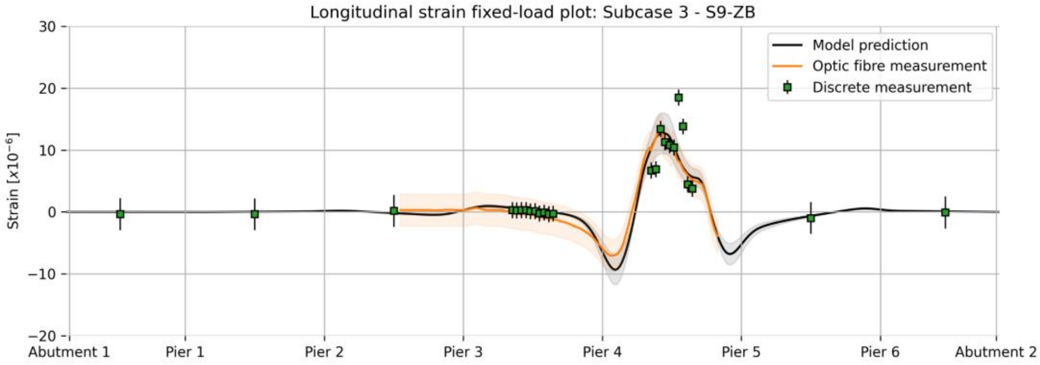
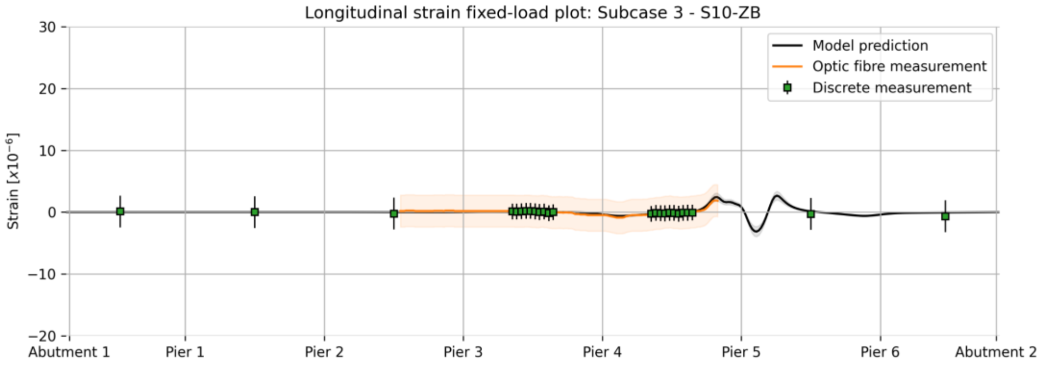


Figure 79: Longitudinal strain diagrams at centreline of Truck 1 lane (deck bottom fibre) for load cases S4-ZB, S5-ZB, S6-ZB and S7-ZB. The model predictions are obtained with the posterior median Sub-case 3 parameters.







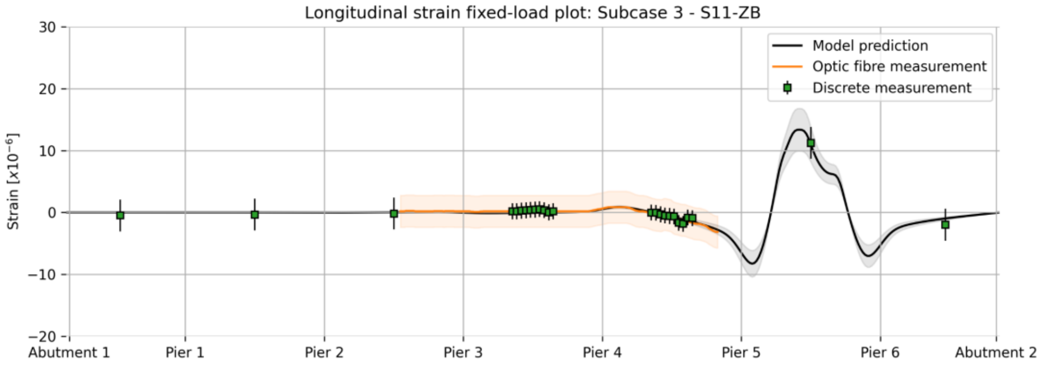


Figure 80: Longitudinal strain diagrams at centreline of Truck 1 lane (deck bottom fibre) for load cases S8-ZB, S9-ZB, S10-ZB and S11-ZB. The model predictions are obtained with the posterior median Sub-case 3 parameters.

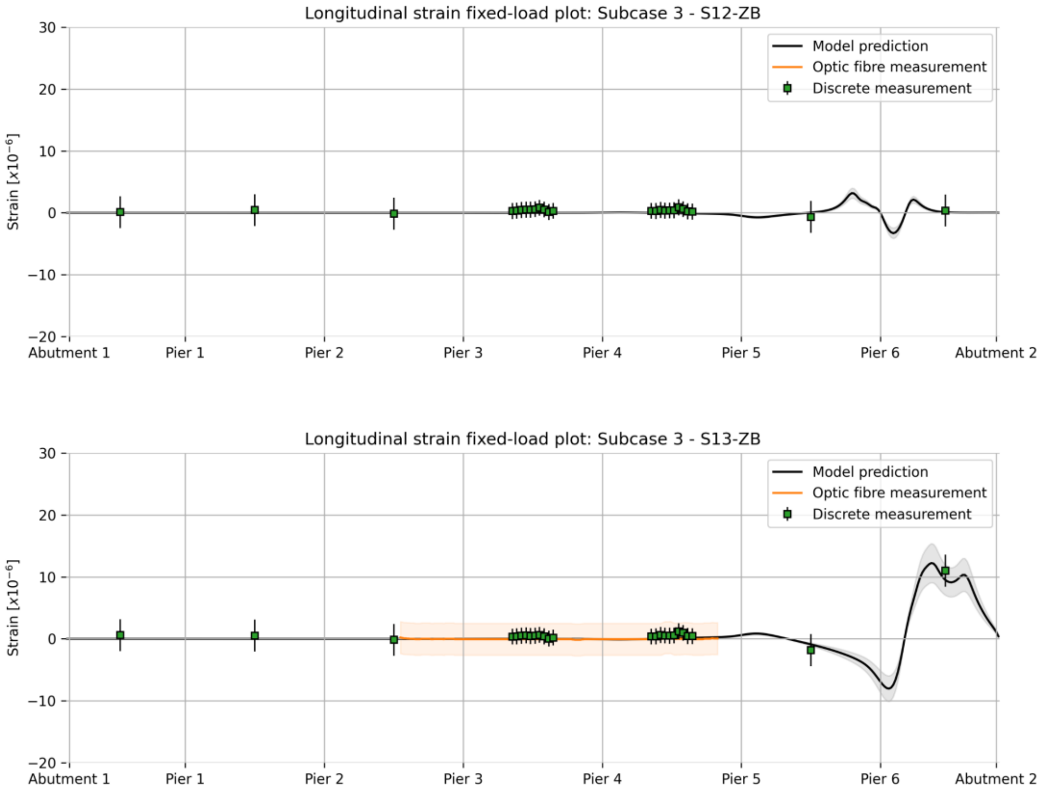


Figure 81: Longitudinal strain diagrams at centreline of Truck 1 lane (deck bottom fibre) for load cases S12-ZB and S13-ZB. The model predictions are obtained with the posterior median Sub-case 3 parameters.

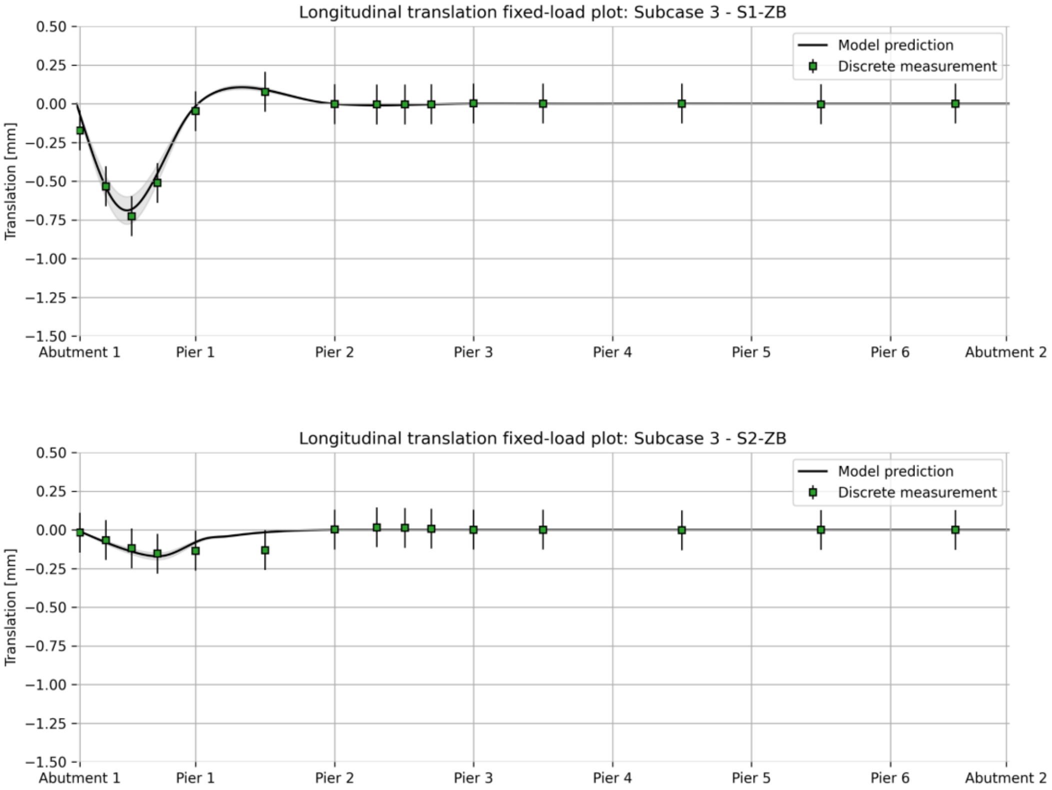
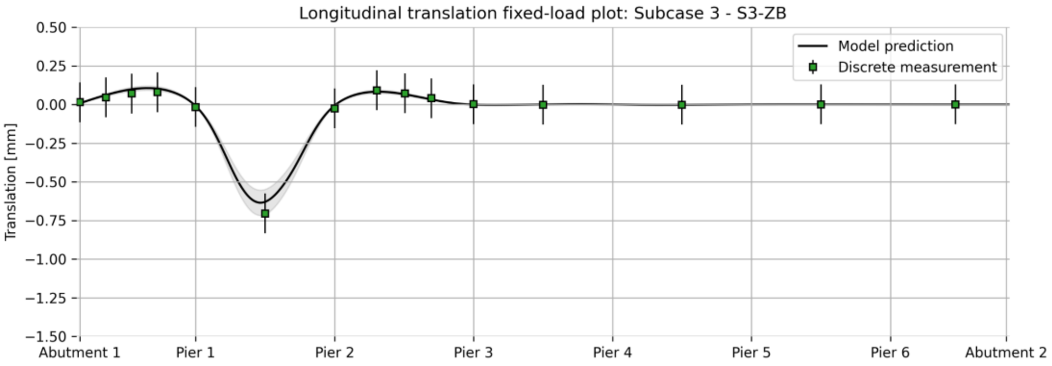
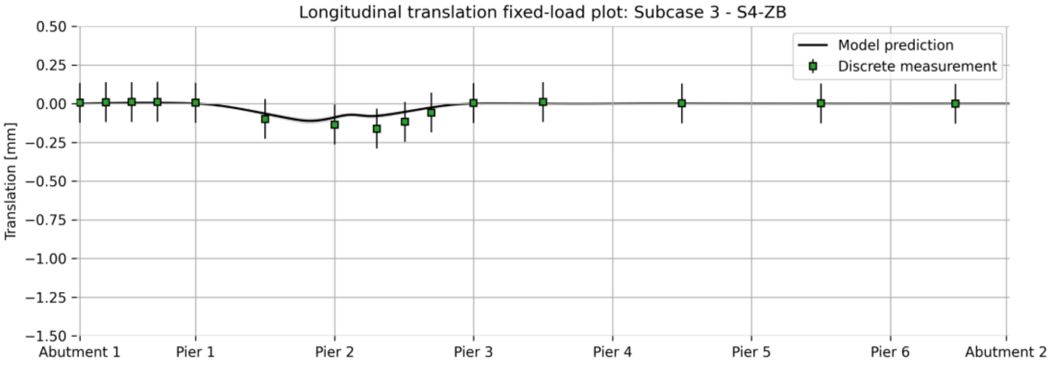
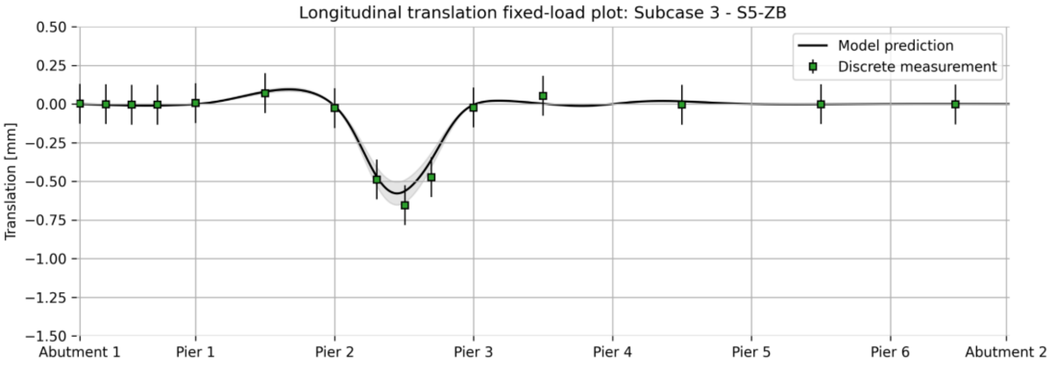


Figure 82: Vertical translation diagrams at centreline of Truck 1 lane for load cases S1-ZB and S2-ZB. The model predictions are obtained with the posterior median Sub-case 3 parameters.







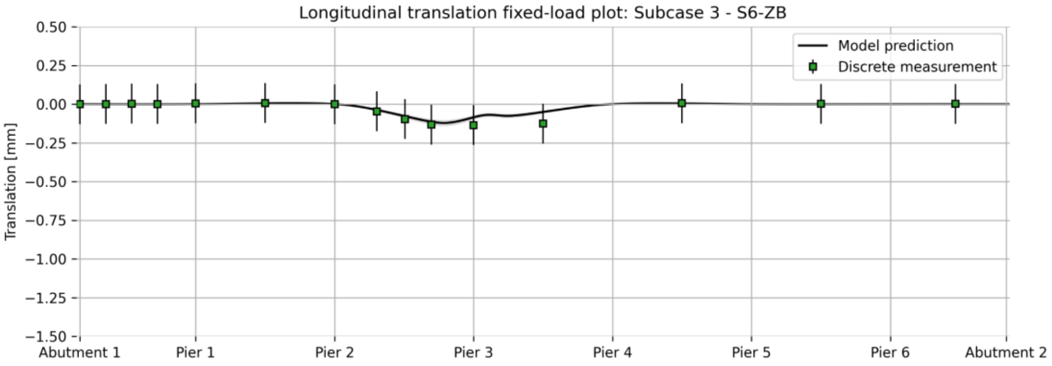
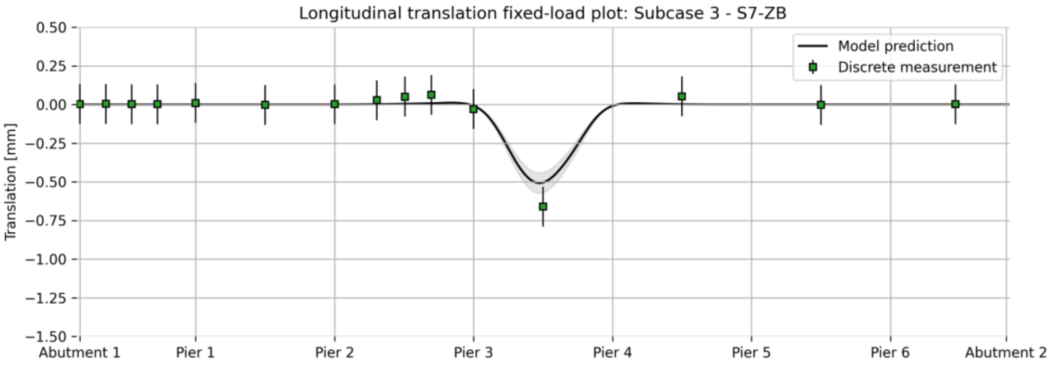
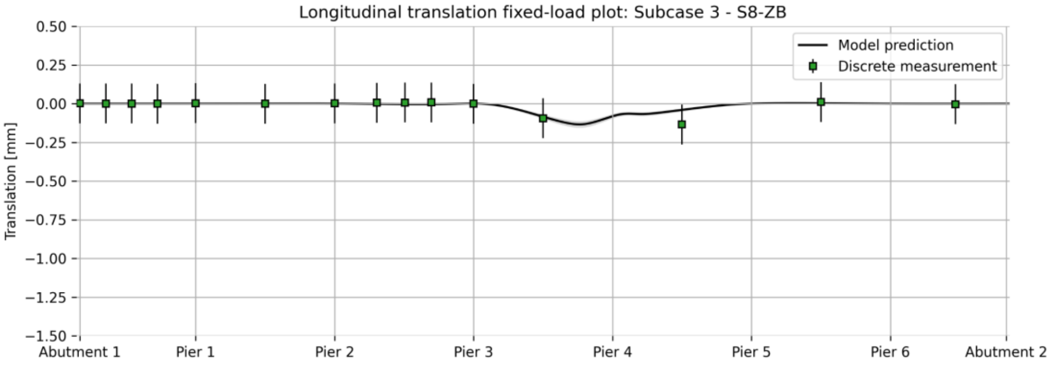
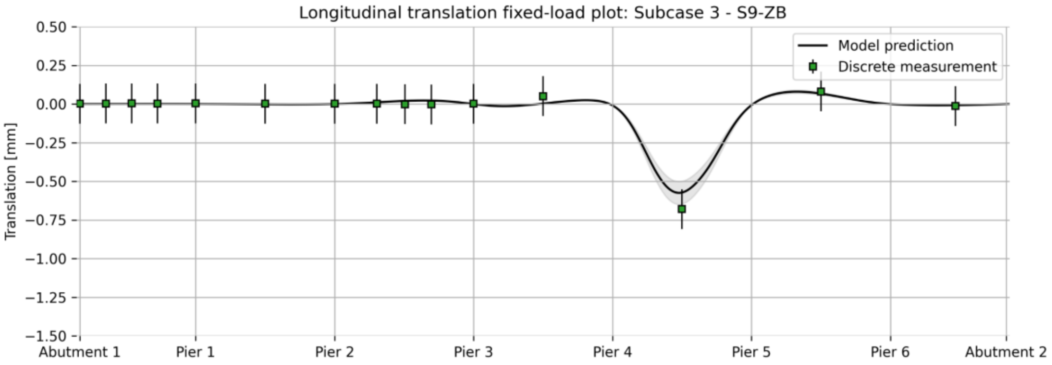


Figure 83: Vertical translation diagrams at centreline of Truck 1 lane for load cases S3-ZB, S4-ZB, S5-ZB and S6-ZB. The model predictions are obtained with the posterior median Sub-case 3 parameters.







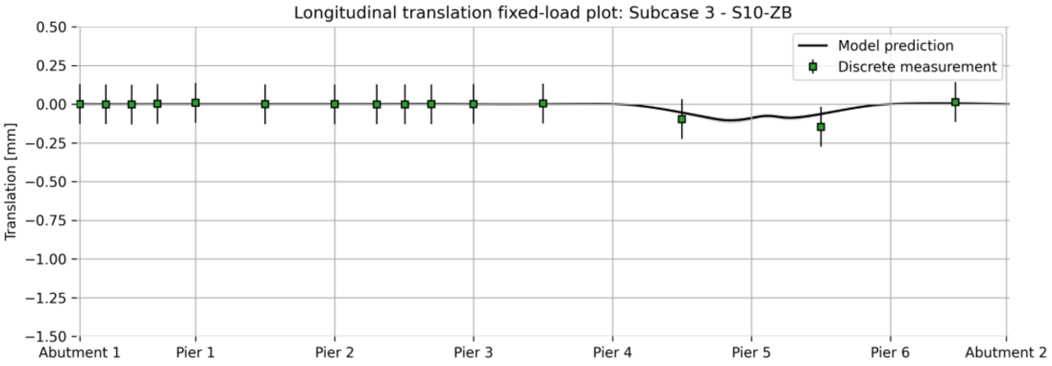


Figure 84: Vertical translation diagrams at centreline of Truck 1 lane for load cases S7-ZB, S8-ZB, S9-ZB and S10-ZB. The model predictions are obtained with the posterior median Sub-case 3 parameters.

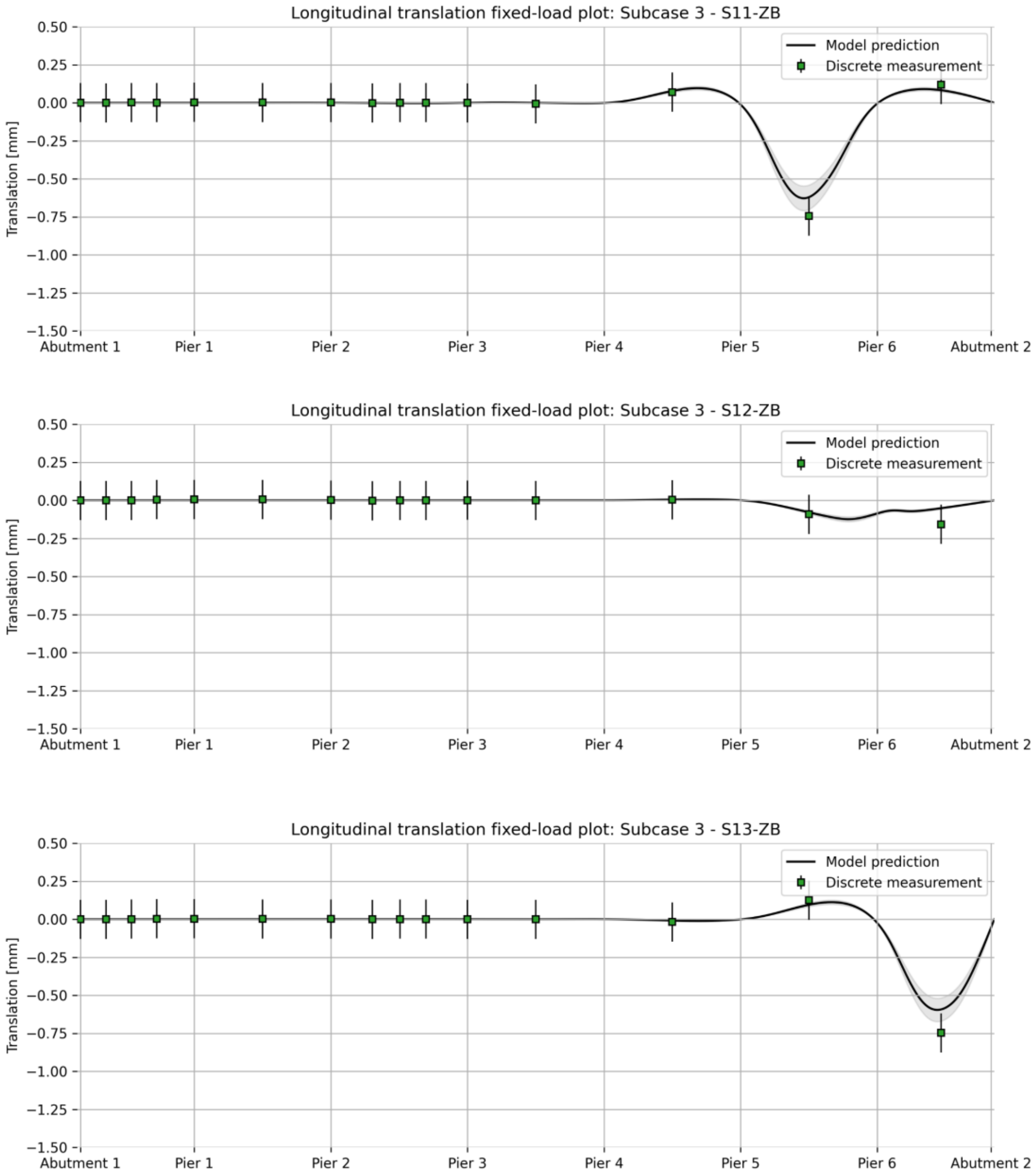


Figure 85: Vertical translation diagrams at centreline of Truck 1 lane for load cases S11-ZB, S12-ZB and S13-ZB. The model predictions are obtained with the posterior median Sub-case 3 parameters.