Bayesian system identification for structures considering spatial and temporal dependencies

Ioannis-Christoforos Koune

TU Delft TNO NTNU

Bayesian system identification for structures considering spatial and temporal dependencies

by

Ioannis-Christoforos Koune

This report is submitted as partial fulfilment of the requirements to obtain the degrees of
Master of Science in Offshore Engineering at Delft University of Technology
&
Master of Science in Wind Energy Technology at Norwegian University of Science and Technology,
under the **European Wind Energy Master** programme.

To be defended publicly on July 5^{th}, 2021 at 15:00 PM.
Officially handed in on June 30^{th}, 2021.

Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology
Department of Marine Technology, Norwegian University of Science and Technology

Project duration:	Sep. 14, 2020 – Jun. 30, 2021	
Thesis committee:	A. Cicirello,	TU Delft, supervisor, chair
	A. A. Nunez Vicencio,	TU Delft
	B. J. Leira,	NTNU
	A. Rozsas,	TNO, supervisor
	A. Slobbe,	TNO, supervisor
	E. Lourens,	TU Delft
	O. Colomes Gene,	TU Delft

An electronic version of this thesis is available at http://repository.tudelft.nl.

TU Delft TNO NTNU

Abstract

Bayesian system identification, including parameter estimation and model selection, is widely used to infer partially known, unobservable parameters of the models of physical systems when measurement data is available. A common assumption in the Bayesian system identification literature is that the discrepancy between model predictions and measurements can be described as independent, identically distributed realizations from a univariate Gaussian distribution. However, the decreasing cost of sensors and monitoring systems leads to more frequent structural measurements in close proximity to each other (e.g. fiber optics and strain gauges). In such cases, dependency in modeling uncertainty could be significant, both in space and time, and the assumption of uncorrelated Gaussian error may lead to inaccurate parameter estimation.

The aim of this thesis is to explore how Bayesian system identification can be feasibly performed using large datasets when spatial and/or temporal dependence might be present and to assess the impact of considering this dependence. A pool of models, each assuming a different correlation structure, is defined and Bayesian inference is performed to obtain posterior distributions of both the physical and probabilistic model parameters. In particular, stress measurements obtained on a twin girder steel road bridge are used to update the parameters of the corresponding finite element model and the parameters of the correlation structure. The results are compared to a reference model where only a small number of measurements of the response peaks are used under the assumption of independence. Nested sampling is utilized to compute the evidence under each model and Bayesian model selection is applied. The question of efficiently performing system identification for large datasets ($N > 10^2$ for temporal dependencies and $N > 10^3$ for combined spatial and temporal dependencies) under multiplicative modeling uncertainty is investigated, and a novel approach for efficiently calculating the exact log-likelihood in the case of spatial and temporal dependencies and additive i.i.d. Gaussian measurement error is derived. Finally, an approximation based on the Fisher information matrix is used to efficiently calculate the information content of measurements.

It is found that the choice of correlation function can significantly affect the posterior distribution of the model prediction uncertainty. Additionally, it is shown that using large datasets and considering dependence makes it possible to perform system identification for a larger number of parameters compared to the reference model. The results of the real-world case study indicate that using measurements from multiple sensors under combined spatial and temporal dependence and additive model prediction error yields reduced uncertainty in the posterior and up to 29% reduction of the posterior predictive credible interval range compared to the reference case. Furthermore, the efficiency of the proposed likelihood evaluation method is assessed. Using this method, exact calculation of the log-likelihood can be performed for $> 10^6$ points in under a second in the case of correlation in one dimension. For combined spatial and temporal correlation it is shown to be approximately 900 times faster than naive evaluation for a 64×64 grid of observations.

The results of the case study indicate that the described approach can be feasibly applied to real-world structures and can potentially improve parameter estimation and reduce prediction uncertainty. These promising findings suggest that further research into the approach could yield substantial improvements over current methods.

Acknowledgements

I would like to extend my gratitude to the following people for their support and contributions to this thesis:

- Arpad Rozsas and Arthur Slobbe, my TNO supervisors who continuously shared their knowledge and provided guidance at every step along the way. I am very thankful for the effort they put, both towards this project and into my own development.
- Alice Cicirello, my TU Delft supervisor for her valuable advice that had a key role in shaping this thesis. Her support and enthusiasm for the project are greatly appreciated.
- Alfredo Nunez Vicencio from TU Delft and Bernt Leira from NTNU, for the valuable feedback and suggestions they provided.

Additionally, I would like to thank Asma Manai for her help with processing the measurements and Geir-Arne Fuglstad for answering my questions on mathematical topics.

Finally, I would like to thank my friends and family, and especially my parents. I am grateful for their unconditional support and constant words of encouragement.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Ensuring the safe and economical operation of civil infrastructure is critical to the well-being of society. The high uncertainty associated with the condition of structures due to the variability of environmental conditions and the unknown deterioration over the design lifetime, combined with the high cost of inspection and maintenance make this a challenging task. These factors highlight the need for accurate assessment of the condition and reliability of infrastructure. Reducing the uncertainty regarding the condition of civil structures and accurately predicting the health of a structure in the future is necessary for better decision making. Structural Health Monitoring (SHM) has the potential to improve the safety of structures, extend their service life time and reduce the cost of maintaining infrastructure (Chen, 2018).

SHM methods for civil structures have recently seen more wide-spread adoption and are now applied to a variety of structures including buildings, bridges, dams, tunnels, offshore oil and gas installations (Brownjohn, 2007) and offshore wind farms (Martinez-Luengo et al., 2016). In order to reduce the need for manual inspection of large structures, which is a cost and labour intensive process that requires expertise and specialized equipment (Kurz et al., 2013), SHM methods based on statistical approaches in combination with physical models (e.g. finite element (FE) models) have seen significant development. These methods use measurements of structural responses obtained from sensors in combination with computational physics models to infer uncertain parameters, calibrate models and provide insight into the structural behaviour. Bayesian statistics offers an approach for combining monitoring data with computational physics models. In Bayesian statistics the problem is cast as a parameter estimation and model selection problem (often referred to as system identification in the SHM literature). Non-directly observed parameters of interest are inferred using directly observed parameters. For example, the rotational stiffness of a support is estimated based on measured deflections.

Obtaining useful measurements for civil structures can pose significant challenges. Since the costs for the equipment, installation and maintenance of monitoring systems can be substantial, and financial resources are limited, the sensor layout must be carefully designed to ensure that accurate and informative data is recorded in a cost-effective manner. Sensor and monitoring technology has seen significant progress in recent years and is capable of providing better accuracy and improved measurement capabilities by utilizing technologies such as fiber optic strain sensors (Barrias et al. (2016), Ye et al. (2014)). These networks provide measurements with high spatial and temporal resolution as large numbers of sensors with high sampling rates are used in the same structure. The larger amount of measurements and higher resolution can potentially improve the accuracy of our predictions, reduce the uncertainty on the inferred system parameters and lead to improved physical models that can more accurately capture structural behaviour. When using measurements from dense sensor layouts, the discrepancies between model prediction and observations are expected to be dependent. Current approaches in Bayesian inference for structures neglect this de-

pendence (Lye et al., 2021). The issue of modeling this dependence, performing Bayesian inference in a computationally efficient manner, and the impact of doing so must therefore be addressed.

1.2 Problem statement

When using closely spaced measurements and model predictions, dependencies may be present in the model prediction errors (Simoen et al., 1998). An illustrative example of the dependencies in the model prediction error for varying distances of two sensors is shown in Figure 1.

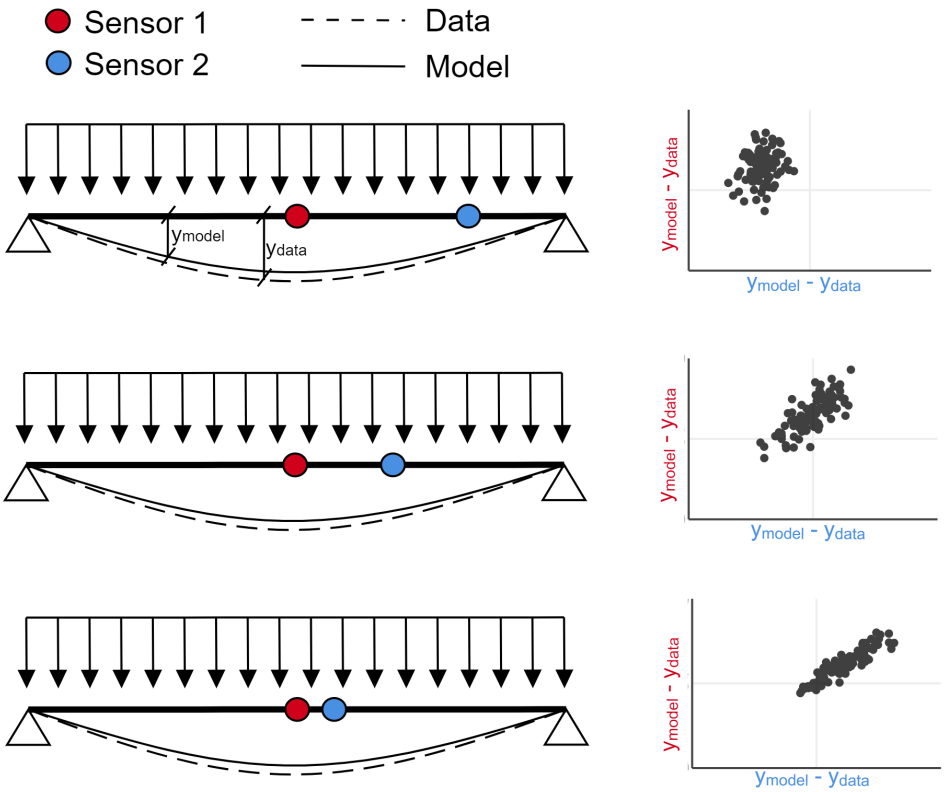


Figure 1: Illustration of dependence in the model prediction error for decreasing distance between two sensors.

Current approaches predominantly neglect dependencies in the measurement error and model prediction error, they assume complete mutual independence (see Section 2.1). However, it is known that some dependency is present over space and time as well. Additionally, through a simple example presented in Appendix D.3 it is shown that assumptions regarding the correlation can have a significant effect on the posterior variance of inferred parameters. Given the lack of a practical approach, i.e. an approach that can be feasibly applied for real-world structures for considering dependencies in Bayesian system identification (see Section 2), the

main research question is defined as follows:

How to perform Bayesian system identification on real-world civil and offshore structures in a practically feasible/manageable way, when spatial and/or temporal dependence might be present?

Considering dependencies in the measurement and model prediction error in Bayesian inference poses significant challenges. This is partially due to the fact that the structure of the dependencies is a-priori unknown. Furthermore, the lack of literature and wide adoption of the independence assumption (see Section 2) in Bayesian system identification for structures results in very limited information being available on the structure of these dependencies. It is also not known to what degree these dependencies are problem specific and how they differ for various structure types and problem domains. Quantifying and modeling these dependencies would be necessary in order to consider them in Bayesian system identification and assess their impact. To address this, the following sub-question is posed:

1. How can we quantify and model the dependencies in the model prediction error?

The computational cost associated with Bayesian system identification is significant (Green et al., 2015), even when dependencies are not taken into account and considering these dependencies further complicates things. Specifically, evaluation of the likelihood for realizations of a high-dimensional multivariate Gaussian distribution requires calculation of the inverse and determinant of the covariance matrix. The computational cost of these operations scales cubically with the number of data points. Combined with the large number of likelihood evaluations ($> 10^3 - 10^4$) typically required by sampling methods (Lye et al., 2021), this makes the computational cost prohibitive for a large number of correlated measurements ($> 10^3$). Furthermore, in order to obtain accurate results, a suitable model must be used. Often, a number of candidate models can describe the system, in which case the available data must be used to determine which of these is the most suitable. Bayesian model selection can be applied to determine the model that best explains the measured data. Calculation of the evidence involves the evaluation of an integral which can be computationally intractable, making it infeasible to apply Bayesian model selection. The applicability of the method in real-world cases is an integral part of this thesis. Therefore, the following sub-question is considered:

2. How can we efficiently perform system identification using large datasets and computationally demanding likelihood functions and evidence?

The amount of information that can be extracted from the measurements when performing system identification is dependent on several factors. These include the number and position of sensors, the inferred parameters, as well as the modeling and measurement uncertainty (Papadimitriou and Lombaert, 2012). Given the difficulty and expense associated with designing and installing sensor networks for structural monitoring, the optimization of the sensor layout is crucial for maximizing the information content of measurements while

minimizing cost. The theory and methods used for sensor layout optimization can also be applied to the similar problem of selecting the most informative measurements, with the aim of reducing problem size and computational effort. To emphasize the importance of the feasibility of the selected approach, the following sub-question is considered:

3. How can we efficiently perform sensor placement optimization?

It is expected that considering these dependencies in the system identification process will have an impact on the resulting parameter estimates and posterior distribution. However, the extent of the impact is not currently known. By quantifying and modeling these dependencies their impact can be studied and the results compared to cases where the dependencies are neglected. This would make it possible to determine the importance of considering dependencies in practical applications, where the aim is to better utilize the available data by including as much of the available measurements as possible. This motivates the following sub-question:

4. What is the impact of considering dependencies in the model prediction error in Bayesian system identification?

1.3 Approach

In this section, a general approach to answer the research questions posed previously is developed as a combination of theory, examples and a real-world use case. An initial literature review is performed to determine relevant methods that may be applied to investigate each question. In addition to reviewing the theoretical background of Bayesian system identification, likelihood evaluation methods and information theory, relevant software packages are also selected. Subsequently, the research questions are explored through simple numerical and analytical examples. These examples are employed to illustrate important concepts and to offer some insights into the feasibility of different methods. Based on these preliminary tests, the most promising methods are further investigated and applied to a real-world use case: a multi-span steel bridge. A high-level overview of this approach is provided in the figure below.

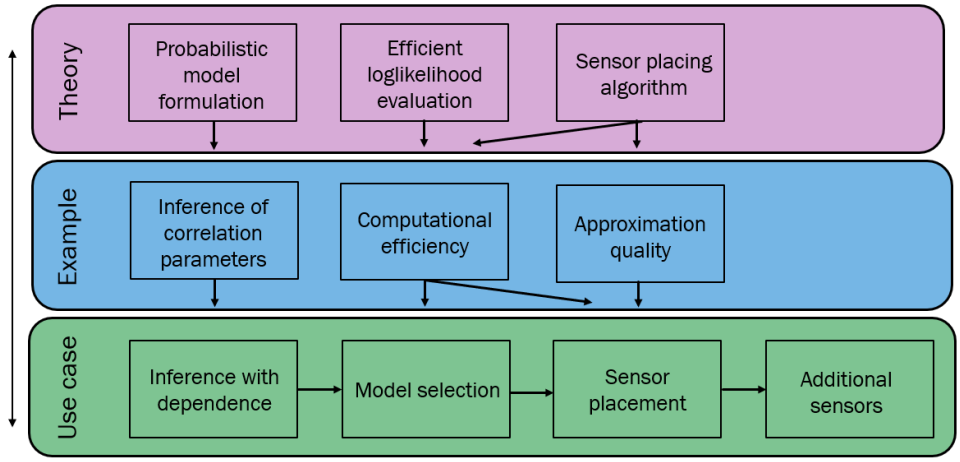


Figure 2: High-level overview of the approach used to address the research questions.

To determine the feasibility and to study the impact of considering dependencies in Bayesian system identification, a mathematical model of the data generating process is defined, assuming a model prediction error and measurement error. A set of probabilistic model classes is formulated with the different models distinguished by the correlation structure considered in the model prediction errors. The probabilistic models are coupled with FE models used to simulate the stress influence lines for selected points in the structure. Following the established system identification literature, a set of unobserved, partially known physical model (structural) parameters are inferred using measured stress influence lines from a real structure. In addition to the structural parameters, a vector of probabilistic model (uncertainty) parameters is also inferred and the posterior of the combined parameter vector is obtained. Bayesian model selection is then applied for the model classes in order to determine the most likely model based on the data. Additionally, the posterior predictive distribution of the stress influence line under each model class is obtained and used to quantify the impact of considering dependencies on the prediction uncertainty. This approach makes it possible to determine the correlation structure, as well as the posterior distribution of the physical and statistical model parameters.

Using this approach, a number of cases are examined. Initially, only a single physical model parameter is considered in order to focus the analysis on the feasibility of performing Bayesian system identification considering dependencies and inferring the correlation structure. A second analysis is performed with additional physical model parameters to improve model fit and assess the feasibility of the approach for larger numbers of parameters, while in the third case the impact of considering multiple sensors is examined. A detailed overview of the analysed cases can be found in Section 6.1

1.4 Scope and limitations

The aim of this thesis is to study the impact of considering spatial and temporal dependencies in the model prediction error in probabilistic system identification, under the Bayesian statistical paradigm, and to detail a feasible approach for doing so in real-world problems. Emphasis is placed on the probabilistic side of the problem, with a focus on computational efficiency when considering datasets with 10^2 - 10^4 correlated measurements and models with up to eight inferred parameters. The damage identification and thorough assessment of the condition of the structure used in the case study is considered to be outside the scope of this thesis. Additionally, only static measurements are considered in order to simplify the analysis and dynamic effects are not taken into account.

Given that the focus is the feasibility and applicability of the approach, only limited discussion is provided regarding the benefits, and no general conclusions are drawn regarding this topic. The aim of the presented approach is not to accurately infer the dependence structure, but rather to determine a simple (in terms of parametrization) probabilistic model that can provide an approximation, to infer the corresponding parameters and to estimate the associated uncertainties. The applicability of the method and conclusions is expected to be independent of the structure and the presented approach can likely be generalized to problems in a variety of engineering disciplines and physical domains.

1.5 Thesis structure

The aim of this introduction chapter is to provide the motivation for the thesis and clearly define the research questions. The approach used to answer these questions, as well as the scope and limitations of the thesis are also described. In Chapter 2, literature related to system identification, efficient likelihood evaluation, sampling and sensor layout optimization is reviewed to determine methods that can be applied to address the research questions. The methods and tools used throughout this thesis, as well as notes on theoretical aspects of these methods are detailed in Chapter 3. A proposal for an efficient likelihood evaluation method is presented in Chapter 4. The necessary mathematical background is established, followed by the derivation and comparison with a conventional approach. The real-world use case is presented in Chapter 5. The analyses and results for the various cases are presented in Chapter 6, followed by conclusions and recommendations in Chapter 7. Additional details, analyses and results not included in the main text but referred to throughout the thesis are provided in Appendices A through D.

1.6 Concluding summary

In this chapter an introduction to the thesis is presented. The motivation behind this thesis is detailed, with the main motivating factor being the need to efficiently perform Bayesian system identification with large datasets and spatially and temporally dependent measurement and model prediction errors, a problem which is not sufficiently addressed in the literature. The issues that must be overcome to achieve this are summarized and

formulated as research questions. The approach used to address the research questions is developed as a combination of theory, examples and a real-world use case. Finally, the scope of the thesis is determined and limitations on the methodology as well as the applicability of the conclusions are defined.

2 Previous work

In this section, literature of select topics is reviewed. For Bayesian system identification, this serves to determine the state of the art and to demonstrate the gap in current approaches that motivates the main research question posed previously. Additionally, methods for efficient likelihood evaluation are reviewed and assessed on the basis of computational efficiency and implementation complexity. Viable approaches are selected for more thorough comparison. Finally, sensor placement optimization literature is reviewed with a focus on computational efficiency. Given the large volume of literature covering these topics, a comprehensive review is outside the scope of this chapter. The aim is to instead provide to indicate the approaches considered, provide references to the most relevant work.

2.1 Bayesian system identification

Bayesian system identification is a probabilistic approach to system identification based on the Bayesian statistical paradigm. This Bayesian system identification framework was established in Beck and Katafygiotis (1998) and has been applied for system identification and damage detection for various types of structures such as bridges (Behmanesh and Moaveni, 2014), rail (Lam et al., 2014) and offshore windfarms (Rogers, 2018). Within this framework, previous knowledge about the system parameters is represented by statistical distributions and combined with measurements to infer the posterior parameter distribution. Bayesian system identification for structures can provide insight into the structural behaviour, make predictions on the output of the system and infer unknown loading conditions and other parameters. Note that the aim of this section is not to review the Bayesian system identification literature, but rather to demonstrate the gap in knowledge that motivates the research questions posed previously. A state of the art review of Bayesian system identification and damage detection for structures can be found in Huang et al. (2019)

Typically in the Bayesian system identification literature, it is assumed that the prediction error is Gaussian white noise, i.e. uncorrelated with zero mean, e.g. in Pasquier and Smith (2015), Chiachío et al. (2015), Astroza et al. (2017), Mthembu et al. (2011). When dependencies are present in the model prediction discrepancies (e.g. when using time series with high sampling rates or spatial data from densely spaced sensors), this assumption can lead to underestimation of the uncertainty and large errors in the estimates and posterior distributions of the inferred parameters. In some studies, e.g. Ebrahimian et al. (2018) and Goller and Schueller (2011) the variance of the model prediction error is included in the vector of inferred parameters, however, dependencies are not considered. In other works, the parameters that define the uncertainty (with or without considering dependencies) are estimated using a subset of the available data (e.g. in Simoen et al. (2015) and Pasquier and Marcotte (2020)), an approach which is not feasible when only a small number of measurements is available. Examples of inference of the uncertainty parameters can be found in applications outside of structural engineering, e.g. in Geostatistics (Diggle and Ribeiro, 2002).

To the authors knowledge, Simoen et al. (2013) is the only work concerning model prediction error correlation in Bayesian system identification for structures, and investigates

the impact of considering dependencies in model prediction discrepancies in the results of Bayesian inference. Bayesian model selection is applied to select a suitable correlation model based on the data and the model and correlation parameters are inferred in a simple linear regression example as well as a reinforced concrete beam example using synthetic modal data. This is done by considering measurement and model prediction error models combined with a structural model. A general set of unobserved model parameters is formed from the corresponding parameters of each model. The study concludes that the prediction error correlation can have a significant influence on the posterior distributions of parameters and must be accounted for and that Bayesian model selection can accurately determine the most probable correlation structure given that sufficient data is available. In this thesis, a similar approach is adopted in order to determine the importance of considering dependency in the model prediction discrepancies and to assess the feasibility of doing so in a real-world case study.

2.2 Efficient likelihood evaluation

Efficient computation of the likelihood is crucial in order to ensure the feasibility of the method detailed throughout this thesis. Typically, evaluation of the joint Gaussian likelihood function requires calculation of the inverse and determinant of the covariance matrix. The complexity and storage requirements of these operations scale poorly with the number of points (see Section 4), making the likelihood costly to evaluate for more than ~ 1000 points (Simpson et al., 2012). Several approaches have been proposed to overcome this issue in the literature of fields such as spatial statistics, Gaussian processes, machine learning and signal processing. Generally, these methods aim to either improve the computational speed or reduce the complexity of the problem by:

- Employing approximations;
- Taking advantage of parallelization and hardware acceleration;
- Exploiting the structure of specific covariance functions;
- Casting the problem into a different mathematical formulation.

A large number of approximate methods have been proposed in the literature. One common approach is covariance tapering Furrer et al. (2012) where elements of the covariance matrix are set to zero, resulting in sparse matrices. In (Litvinenko et al., 2019) a scheme of hierarchical approximations is applied to approximate a dense covariance matrix using the $\mathcal{H}$-matrix format, resulting in reduced memory requirements and computational cost. In the widely used Vecchia approximation (Vecchia, 1988), the likelihood is factored into a series of conditional distributions to improve computational efficiency. This approach was later generalized into the Vecchia framework (Katzfuss and Guinness, 2019) which also includes other popular approximate methods such as the composite likelihood Varin (2008).

Other approaches utilizing numerical approximation of Gaussian Markovian random fields (GMRF) (Rue and Held, 2005) have been applied to spatial statistics (Simpson et al., 2012).

Taking advantage of the Markov property of certain kernel functions (see Section 3.3.1), this approach significantly reduces the computational cost associated with matrix operations in the likelihood calculation and makes it possible to derive analytical expressions of linear computational complexity for arbitrarily large covariance matrices for time-series data (Pasquier and Marcotte, 2020). The approach presented in that work is simple to implement and makes inference feasible for datasets with $> 10^6$ points. It is however limited to specific covariance functions, one-dimensional problems and regularly spaced observations, making it impractical for the purpose of this thesis. This is further discussed in Section 4, where the Markov property is used to derive a novel approach for efficient likelihood evaluation that is applicable to spatio-temporal problems on an irregular grid.

The GMRF approach is closely related to the field of stochastic partial differential equations (SPDE's) (Sarkka and Solin, 2019). These methods utilize the connection between Gaussian processes and stochastic differential equations to reformulate the problem such that it can be solved by numerical methods for partial differential equations. Other similar approaches achieve a significant reduction in computational complexity (Fuglstad et al., 2014) for spatial and spatio-temporal models by deriving a state-space representation for different covariance functions (see Solin (2016)). Despite the advantages they offer, these methods are challenging to implement and can not be readily applied to the probabilistic models used in this thesis

Finally, a number of approaches have been developed that take advantage of parallelism and GPU acceleration to speed up inference and machine learning with Gaussian processes for large datasets. Such an approach is presented in (Gardner et al., 2018), where Blackbox Matrix-Matrix multiplication is used to obtain improved performance for Gaussian processes on GPUs and further developed in Wang et al. (2019) where it is parallelized over multiple GPUs resulting in $O(N)$ memory requirements. The complexity of this technique scales quadratically with the number of points, allowing for inference to be performed on large datasets with no constraints on the covariance structure or the spacing of the observations. In Abdulah et al. (2018), a software package for high performance computing utilizing parallel architectures and distributed memory is developed and used for maximum likelihood estimation with exact loglikelihood evaluation for large Geostatistics problems, demonstrating the possible performance gain from utilizing modern hardware.

2.3 Concluding summary

A review of recent advancements and applications of Bayesian system identification highlights the relevance of the main research question posed in Section 1.2. The Gaussian white noise structure of the measurement and model prediction error is prevalent in the literature, while studies that consider dependency do so by using a subset of the available data to determine its structure. The former approach is applicable when a sufficient number of samples is available, making it infeasible when data is limited. In Simoen et al. (2013), Bayesian system identification is applied to infer the structure and parameters of the dependency for a reinforced concrete beam model using synthetic data. The lack of similar applications to

real-world cases constitutes a gap in the literature that is addressed in this thesis.

Additionally, relevant approaches for efficient likelihood evaluation are reviewed and their advantages and disadvantages are briefly examined with a focus on computational cost, ease of implementation and the degree to which they can be generalized. It is determined that no single method satisfies all three criteria and a trade-off must be made. Based on the work of Pasquier and Marcotte (2020), an approach utilizing the sparse inverse structure for a class of covariance functions is derived in Section 4.

3 Review of methods for Bayesian system identification and sensor placement

The theoretical background of the methods and tools used throughout this thesis to perform Bayesian system identification is established in this section. A set of models of the real physical process being studied is formulated as a combination of a deterministic physical model and a number of probabilistic models. The approach used to combine models, measurements and prior knowledge using Bayesian statistics to obtain the likelihood of different models, infer uncertain parameters and predict future observations is detailed. Finally, a summary of the information theoretical approach used to perform measurement selection is provided.

3.1 Bayesian parameter estimation, model selection and posterior predictive

3.1.1 Continuous Bayes theorem

The Bayes theorem of conditional probability for continuous random variables can be written as (Gelman et al., 2013):

$$p(\boldsymbol{\theta}|\boldsymbol{y},\mathcal{M}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta},\mathcal{M})p(\boldsymbol{\theta}|\mathcal{M})}{\int_\Theta p(\boldsymbol{y}|\boldsymbol{\theta},\mathcal{M})p(\boldsymbol{\theta}|\mathcal{M})d\boldsymbol{\theta}} \tag{1}$$

where:

- $\boldsymbol{\theta}$ is a vector of uncertain parameters;
- $\boldsymbol{y}$ a vector of measurements;
- $\mathcal{M}$ denotes the model;
- $p(\boldsymbol{\theta}|\boldsymbol{y},\mathcal{M})$ is the posterior distribution;
- $p(\boldsymbol{y}|\boldsymbol{\theta},\mathcal{M})$ is the likelihood;
- $p(\boldsymbol{\theta}|\mathcal{M})$ is the prior.

It can be seen that $p(\boldsymbol{\theta}|\boldsymbol{y},\mathcal{M})$ describes the posterior distribution of the model parameter set $\boldsymbol{\theta}$ conditional on measurements $\boldsymbol{y}$ under model $\mathcal{M}$[1]. The likelihood term, also denoted as $L(\boldsymbol{\theta})$, gives the probability of observing $\boldsymbol{y}$ given parameters $\boldsymbol{\theta}$, while the prior distribution describes our previous knowledge or belief (meaning an assumption based on engineering judgement in this context) regarding the distribution of the parameters. Finally, the denominator in the right hand side is known as the evidence, or marginal likelihood and gives the likelihood of obtaining the measurements conditional on the model $\mathcal{M}$, which is defined as a combination of a physical and a statistical model. This term also acts as a normalizing constant that ensures the posterior distribution is a proper probability density function, i.e. integrates to unity. The evidence will henceforth be also denoted as $\mathcal{Z} = p(\boldsymbol{y}|\mathcal{M}_i)$. In most

[1]The dependence on the model $\mathcal{M}$ is typically omitted from the notation.

practical applications this integral is high dimensional (see e.g. Lye et al. (2021)) and can typically not be computed by numerical integration. Obtaining the evidence is also necessary for performing Bayesian model selection.

In many engineering applications, the available information on a system is limited. The Bayesian approach defines a statistical framework for applying probabilistic methods in such cases. Using the Bayes theorem, the distributions of uncertain parameters can be updated as new information is obtained, enabling the combination of prior knowledge with new data. To achieved this, the Bayesian method considers the distribution parameters of uncertain variables to be uncertain variables themselves. This makes it possible to combine different sources of uncertainty in a formal way (Ang and Tang, 2007).

3.1.2 Bayesian model selection

In addition to the data, the posterior distribution of the parameters is dependent on the model $\mathcal{M}$. The model is typically not known a priori. In practice a class of models $\boldsymbol{\mathcal{M}}$ is often defined and inference is performed conditional on each model $\mathcal{M}_i$. The Bayes rule can then be applied to select the most likely model. In Hoeting et al. (1999), the following equation is provided for performing Bayesian model selection:

$$p(\mathcal{M}_i|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\mathcal{M}_i)p(\mathcal{M}_i)}{\sum_{i=1}^{K} p(\boldsymbol{y}|\mathcal{M}_i)p(\mathcal{M}_i)}, \tag{2}$$

where:

- $p(\mathcal{M}_i|\boldsymbol{y})$ is the posterior probability of model i;
- $p(\boldsymbol{y}|\mathcal{M}_i)$ is the likelihood of that model;
- $p(\mathcal{M}_i)$ is the prior probability of model i.

It should be emphasized that a high posterior model probability does not necessarily indicate that a particular model provides a good fit with the data since model probabilities are conditioned on the pool of candidate models $\boldsymbol{\mathcal{M}}$. Therefore, a high posterior model probability can only be interpreted as a particular model being more likely relative to the other models in a given class. The likelihood term $p(\boldsymbol{y}|\mathcal{M}_i)$ is the evidence under model i and can be obtained by evaluating the integral in Equation 1. This integral is typically high-dimensional and requires a large number of evaluations of the physical model to determine, meaning that it is computationally intractable in most applications. It is therefore evaluated using numerical methods.

Given a class of models, selecting the model that best fits the data can straightforwardly be achieved by selecting the model that minimizes a particular error metric between measurements and model outputs. However, simply choosing the model that best fits the data could potentially lead to overfitting. More complicated models would tend to fit the data

best, making them the most likely in this approach even if the added complexity provides a negligible benefit. An advantage of Bayesian model selection is that it automatically enforces model parsimony, also known as Occam's razor as discussed in MacKay (2003) and Beck and Yuen (2004), penalizing overly complex models.

3.1.3 Bayes factor

The posterior probability of a model $\mathrm{p}(\mathcal{M}_i|\boldsymbol{y})$ is often not sufficient for comparing different models as it does not provide an easily interpretable measure of the evidence in favour of a particular model. Instead, the Bayes factor given in Equation 3 can be used to compare the plausibility between models $\mathcal{M}_i$ in model class $\mathcal{M}$. For two models $\mathcal{M}_1$, $\mathcal{M}_2$ belonging to a class of models $\boldsymbol{\mathcal{M}}$, the Bayes factor is a measure of the relative plausibility of the two models.

$$K = \frac{p(\mathcal{M}_1|\boldsymbol{y})}{p(\mathcal{M}_2|\boldsymbol{y})} \cdot \frac{p(\mathcal{M}_2)}{p(\mathcal{M}_1)} \tag{3}$$

An advantage of using the Bayes factor over the posterior model probabilities for model selection is that it can be readily interpreted to indicate the support of one model over another and thus offers a practical means of comparing different models. Various interpretations have been suggested in the literature, e.g. in Jeffreys (2003) given in the table below.

Table 1: Interpretation of the Bayes factor from Jeffreys (2003).

K	Strength of evidence
$<10^0$	Negative
10^0 to $10^{1/2}$	Barely worth mentioning
$10^{1/2}$ to 10^1	Substantial
10^1 to $10^{3/2}$	Strong
$10^{3/2}$ to 10^2	Very strong
$>10^2$	Decisive

3.1.4 Posterior predictive distribution

Bayesian system identification can be used to obtain point estimates and posterior distributions of uncertain parameters using physical models and measurement data. However, directly using estimates of the inferred parameters to make predictions would result in underestimation of the uncertainty and overly confident predictions. This is due to the fact that using point estimates to make predictions disregards the uncertainties in the inferred parameters resulting from lack of data. In contrast, the posterior predictive is a distribution of possible future observations conditioned on past observations and can be obtained as (Gelman et al., 2013):

$$p(\tilde{\boldsymbol{y}}|\boldsymbol{y}) = \int_{\Theta} p(\tilde{\boldsymbol{y}}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}|\boldsymbol{y}) \cdot \mathrm{d}\boldsymbol{\theta}, \tag{4}$$

where:

- $\tilde{\boldsymbol{y}}$ is a vector of possible future observations
- $\boldsymbol{y}$ is a vector of observations

The posterior predictive takes into account the combined uncertainty from all sources (e.g. modeling, measurement and parameter uncertainty) and can therefore correctly estimate the probability of future observations.

3.2 Data generating process

In order to perform system identification, the likelihood function in Equation 1 is formulated based on the combination of a probabilistic model and a deterministic physical model. This coupled probabilistic-physical model is used to represent the process that generates the measurements, referred to as the *data generating process*. Details on the deterministic physical model are provided in Section 5.3. The probabilistic model is used to represent the uncertainties that are inherent when using a model to describe a physical system.

The physical models used to represent structures can never be considered to be perfect representations of reality. There is always some degree of uncertainty present in regards to the model of any physical system. The following sources of uncertainty are considered:

- Measurement uncertainty
- Physical model uncertainty

Measurement uncertainty refers to the discrepancies between the measured response quantities and the true system response, caused by the combined influence of sensor errors and environmental noise (Kennedy and O'Hagan, 2001). Modeling uncertainty can contain several components and refers to the error between reality and the models used to represent it. These errors arise due to physical simplifications used to derive system responses and numerical approximations.

The data generating processes described by Equation 5 and Equation 7 are obtained by considering the discrepancies between the deterministic model output and the real system response, a process referred to as *stochastic embedding* in Beck (2010). Two models are formulated, distinguished by the type of model prediction uncertainty.

3.2.1 Additive modeling uncertainty

The data generating process assuming additive modeling uncertainty is considered, given in Equation 5. In the following, capital letters are used to represent random variables.

$$\boldsymbol{Y}_{\text{real}} = \boldsymbol{y}_{\text{model}}(\boldsymbol{\theta}_{\text{s}}) + \boldsymbol{E}_{\text{model}}(\boldsymbol{\theta}_{\text{c}}) + \boldsymbol{E}_{\text{meas}}, \tag{5}$$

where:

- $\boldsymbol{Y}_{\text{real}}$ is a vector of observation random variables;
- $\boldsymbol{y}_{\text{model}}$ is a vector valued physical model function;
- $\boldsymbol{E}_{\text{model}}$ is a vector of physical model prediction error random variables;
- $\boldsymbol{E}_{\text{meas}}$ is a vector of measurement error random variables;
- $\boldsymbol{\theta}_{\text{s}}$ is a vector of physical model parameters;
- $\boldsymbol{\theta}_{\text{c}}$ is a vector of probabilistic model parameters.

In this formulation, the magnitude of the model prediction error is considered independent of the model output. The measurement uncertainties are taken as i.i.d. Gaussian random variables, distributed as $\boldsymbol{E}_{\text{meas}} \sim N(0, \boldsymbol{\Sigma}_{\text{meas}})$. The assumption of Gaussian white noise for the measurement uncertainty is prevalent in the literature and is commonly used in Bayesian system identification for structures (see Section 2), stemming from the fact that measurement noise can be considered as a sum of a large number of random variables and therefore the central limit theorem justifies using an i.i.d. Gaussian distribution to model it. This results in the following description of the combined probabilistic-physical model:

$$\boldsymbol{Y}_{\text{real}} \sim N(\boldsymbol{y}_{\text{model}}(\boldsymbol{\theta}_{\text{s}}), \boldsymbol{\Sigma}_{\text{model}}(\boldsymbol{\theta}_{\text{c}}) + \boldsymbol{\Sigma}_{\text{meas}}) \tag{6}$$

It is noted that this formulation assumes that the modeling uncertainty is independent of the magnitude of the model output. This can lead to unrealistic results, e.g. in the case of modeling uncertainty close to points in the physical model where the boundary conditions ensure zero model output. The lower modeling uncertainty in this case is not taken into account when additive modeling uncertainty is considered.

3.2.2 Multiplicative modeling uncertainty

A probabilistic model with the assumption of multiplicative modeling uncertainty is also considered. In this case the model has the form:

$$\boldsymbol{Y}_{\text{real}} = \boldsymbol{K}_{\text{model}}(\boldsymbol{\theta}_{\text{c}}) \cdot \boldsymbol{y}_{\text{model}}(\boldsymbol{\theta}_{\text{s}}) + \boldsymbol{E}_{\text{meas}}, \tag{7}$$

where the model predictions are multiplied by a factor $\boldsymbol{K}_{\text{model}}$, expressing the discrepancy between model prediction and reality. A correlated multivariate normal distribution with a mean of 1.0 is assumed for the $\boldsymbol{K}_{\text{model}}$ as shown in Equation 8. The formulation of the covariance $\boldsymbol{\Sigma}_{\text{model}}$ is discussed in Section 3.3.2.

$$\boldsymbol{K}_{\text{model}}(\boldsymbol{\theta}_{\text{c}}) \sim N(1.0, \boldsymbol{\Sigma}_{\text{model}}(\boldsymbol{\theta}_{\text{c}})) \tag{8}$$

The assumption of a Gaussian distribution for $\boldsymbol{K}_{\text{model}}$ is made primarily for simplicity and computational convenience. The impact of this assumption is not further examined. An additional i.i.d. Gaussian noise term $\boldsymbol{E}_{\text{meas}}$ is also considered to account for the measurement uncertainty with a mean of zero and a diagonal covariance matrix:

$$\boldsymbol{\Sigma}_{\text{meas}} = \sigma^2_{\text{meas}} \boldsymbol{I} \tag{9}$$

In this model formulation it is assumed that the uncertainty in the physical model prediction scales with the magnitude of the model output. This assumption is prevalent in the structural reliability literature (Cervenka et al. (2018) and Sykorka et al. (2018)), and can potentially yield more realistic results. As an example, the modeling uncertainty in the area of a boundary condition that imposes a zero response is considered. Under the multiplicative assumption the lower modeling uncertainty near the boundary, imposed by the boundary condition, is taken into account. In contrast, under the additive assumption the uncertainty near a boundary with zero response is the same as for all other positions. The real system response is modeled as:

$$\boldsymbol{Y}_{\text{real}} \sim N(\boldsymbol{y}_{\text{model}}(\boldsymbol{\theta}_{\text{s}}), \boldsymbol{\Sigma}_{\text{phys}}(\boldsymbol{\theta}) + \boldsymbol{\Sigma}_{\text{meas}}), \tag{10}$$

where $\boldsymbol{\Sigma}_{\text{phys}}$ is obtained as $\boldsymbol{Y} \cdot \boldsymbol{\Sigma}_{\text{model}}(\boldsymbol{\theta}_{\text{c}}) \cdot \boldsymbol{Y}$ with $\boldsymbol{Y} = \text{diag}(\boldsymbol{y}_{\text{model}}(\boldsymbol{\theta}_{\text{s}}))$.

3.2.3 Likelihood function

Assuming a joint normal distribution for the discrepancies between the physical model output and the true system response, the following expression is obtained for the likelihood:

$$L(\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}(\boldsymbol{\theta}_c)|}} \exp\left[-\frac{1}{2}(\boldsymbol{y}_{\text{meas}} - \boldsymbol{y}_{\text{model}}(\boldsymbol{\theta}_s))^T \boldsymbol{\Sigma}(\boldsymbol{\theta}_c)^{-1} (\boldsymbol{y}_{\text{meas}} - \boldsymbol{y}_{\text{model}}(\boldsymbol{\theta}_s))\right] \tag{11}$$

In the previous equation, $\boldsymbol{y}_{meas}$ denotes a particular realization of the vector of measurement random variables and $|\boldsymbol{\Sigma}|$ and $\boldsymbol{\Sigma}^{-1}$ are respectively the determinant and inverse of the diagonal covariance matrix. Calculation of those terms is computationally demanding for large N. This poses a significant challenge, given that efficient evaluation of the log-likelihood is critical for feasibly performing inference for large numbers of measurements. This is discussed in detail in Section 4.

3.3 Correlations

3.3.1 Correlation functions

Correlation functions[2] are positive definite functions of two Euclidean vectors $k(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\theta}_c)$ (Duvenaud, 2014) that describe the correlation between points $\boldsymbol{x}$ and $\boldsymbol{x}'$. The correlation functions considered in this thesis are summarized in Table 2.

[2]Also referred to as kernels or kernel functions in the literature and throughout this report.

Table 2: List of correlation functions and corresponding uncertainty parameters.

Kernel	$k(\boldsymbol{x}, \boldsymbol{x}')$	Parameters $\boldsymbol{\theta}_c$	
I.i.d.	I	-	-
Radial Basis	$\exp\left(-\frac{\|\boldsymbol{x},\boldsymbol{x}'\|^2}{2l_{\mathrm{corr}}^2}\right)$	l_{corr}	-
Rational Quadratic	$\left(1+\frac{\|\boldsymbol{x},\boldsymbol{x}'\|^2}{2\alpha l_{\mathrm{corr}}^2}\right)^{-\alpha}$	l_{corr}	α
Damped Cosine	$\exp\left(-\frac{\|\boldsymbol{x},\boldsymbol{x}'\|}{l_{\mathrm{corr}}}\right)\cdot\cos\left(w_n\|\boldsymbol{x},\boldsymbol{x}'\|\right)$	l_{corr}	w_{n}
Matern $\nu = 1.5$	$\frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{\sqrt{2\nu}}{l_{\mathrm{corr}}}\|\boldsymbol{x},\boldsymbol{x}'\|\right)^{\nu}K_{\nu}\cdot\left(\left(\frac{\sqrt{2\nu}}{l_{\mathrm{corr}}}\|\boldsymbol{x},\boldsymbol{x}'\|\right)\right)$	l_{corr}	-
Exponential	$\exp\left(\frac{\|\boldsymbol{x},\boldsymbol{x}'\|}{l_{\mathrm{corr}}}\right)$	l_{corr}	-

The symbols and variables used in the previous table are defined as follows:

- $\|\cdot\|$ is the Euclidean distance;
- l_{corr} is the correlation length parameter;
- α is the scale mixture parameter;
- w_{n} is the wave number;
- ν is the smoothness parameter.

The shape of a correlation function is determined by the parameter vector $\boldsymbol{\theta}_c$. The correlation length, or lengthscale parameter l_{corr} is of particular importance as it defines the distance over which a pair of points will exhibit correlation or the length over which the correlation decays. This is illustrated in Figure 4, where realizations of a Gaussian random field with exponential correlation for increasing correlation length are shown.

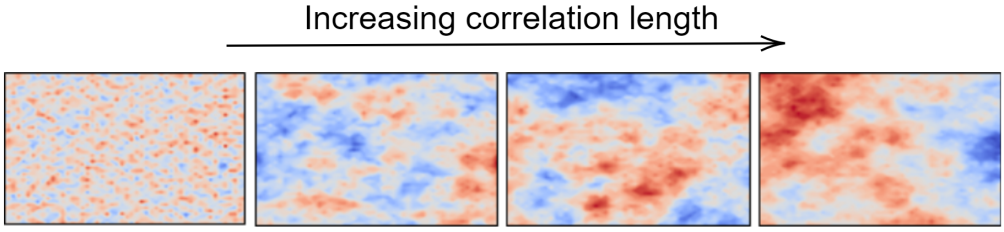


Figure 3: Realizations of a Gaussian random field with exponential correlation for increasing correlation length.

This easily interpretable physical meaning makes the correlation length a parameter of interest. Another such parameter is the wave number w_n, which determines the oscillation cycles per unit length for the damped cosine kernel. In contrast, other kernel parameters can not be readily interpreted and are therefore considered nuisance parameters. This includes the scale mixture parameter α and the smoothness parameter ν. The latter is predefined as it is typically poorly identifiable (Hu et al., 2013). Finally, it should be noted that the exponential kernel is a special case of the Matern kernel class for $\nu = 1/2$, while the radial basis kernel can be obtained from the rational quadratic for $\alpha \to \infty$ (Rasmussen and Williams, 2006). An overview of the correlation functions listed in Table 2 and their behaviour for different parameter values is provided in the figure below.

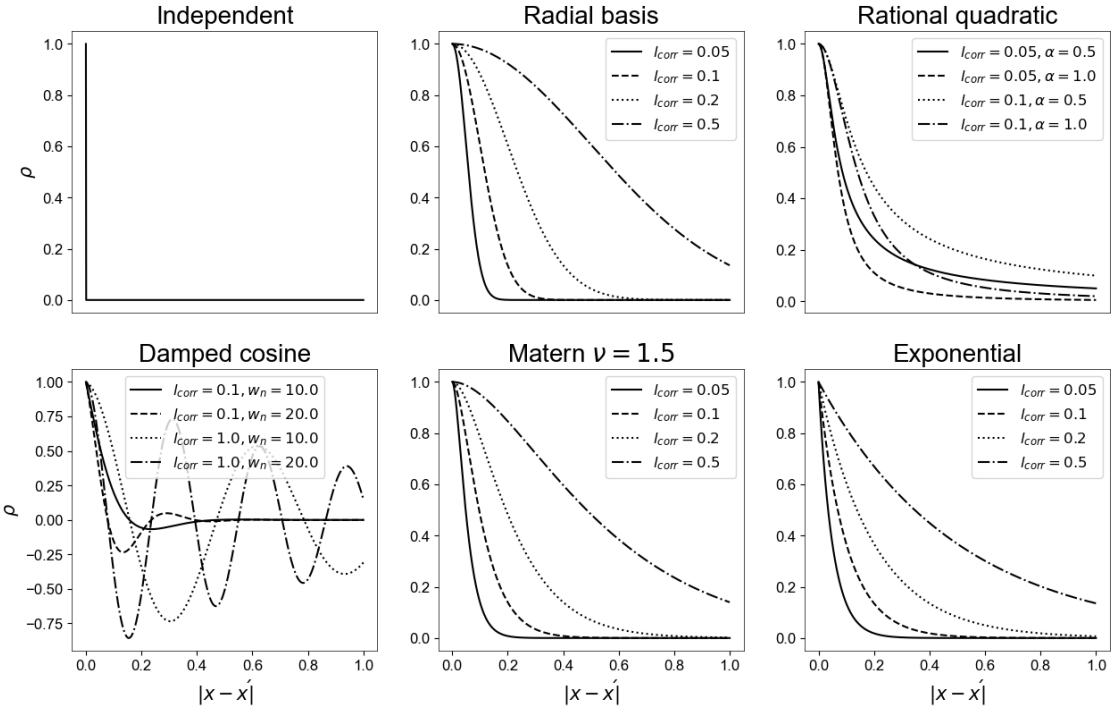


Figure 4: Correlation coefficient as a function of distance for varying parameter values of each considered kernel.

The process of choosing these functions was based on two main factors:

- Obtaining a set of functions that can capture a wide range of correlation structures.
- Obtaining models with different numbers of parameters to investigate the effect of model complexity on Bayesian model selection

3.3.2 Covariance

The residuals between measurements and model predictions are considered as a random process or random field in the probabilistic model formulation presented in Section 3.2. The correlations in the modeling uncertainties between two points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ can be described in terms of a kernel function (see Section 3.3.1):

$$\rho_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{\theta}_\mathrm{c}), \tag{12}$$

where $k(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{\theta}_\mathrm{c})$ is parametrized by the set of statistical model parameters $\boldsymbol{\theta}_\mathrm{c}$. The vector $\boldsymbol{x}_i$ denotes the coordinates of an observation in Euclidean space (e.g. in the case of inference with observations from multiple sensors presented in 6.4, $\boldsymbol{x}$ is a vector describing the sensor location $x_{i,1}$ and load position $x_{i,2}$ in the longitudinal axis of a bridge):

$$\boldsymbol{x}_i = [x_{i,1}, x_{i,2}, ..., x_{i,D}] \tag{13}$$

where D is the number of dimensions. The correlation function is multiplied by the standard deviation of the modeling uncertainty to form the covariance function. Calculating the covariance for every pair of points yields the symmetric positive semi-definite covariance matrix shown in Equation 14.

$$\Sigma = \begin{bmatrix} \sigma_1^2 k(\boldsymbol{x_1}, \boldsymbol{x_1}; \boldsymbol{\theta}_\mathrm{c}) & \dots & \sigma_1 \sigma_N k(\boldsymbol{x_1}, \boldsymbol{x_N}; \boldsymbol{\theta}_\mathrm{c}) \\ \vdots & \ddots & \vdots \\ \sigma_N \sigma_1 k(\boldsymbol{x_N}, \boldsymbol{x_1}; \boldsymbol{\theta}_\mathrm{c}) & \dots & \sigma_N^2 k(\boldsymbol{x_N}, \boldsymbol{x_N}; \boldsymbol{\theta}_\mathrm{c}) \end{bmatrix}, \tag{14}$$

Note that the marginal variance parameter σ is included here for clarity but can equivalently be included in the covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$ (see Section 3.3.1).

3.4 Bayesian system identification using nested sampling

Bayesian system identification relies on sampling techniques to obtain the posterior distribution of uncertain parameters. Starting with the introduction of Markov Chain Monte Carlo (MCMC) by Metropolis et al. (1953) and further development by Hastings (1970) several algorithms have been proposed, with the majority of them being geared towards efficiently estimating the posterior. However, standard MCMC techniques can not compute the evidence and are therefore not applicable for the purpose of model selection. Nested sampling is a technique proposed by Skilling (2006) that was primarily developed for calculation of the evidence in Bayesian inference, i.e. evaluating the integral:

$$\mathcal{Z} = \int L \, \mathrm{d}X \tag{15}$$

where $\mathcal{Z}$ is the evidence, $L(\boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood function and $\mathrm{d}X = p(\boldsymbol{\theta})\mathrm{d}\underline{\theta}$ is the prior mass element. Therefore, nested sampling addresses one of the key issues of standard MCMC. In this thesis the Nested sampling approach is chosen over other alternatives such as MCMC as, in addition to calculating the evidence, it can also provide estimates Kullback-Leibler divergence and can deal effectively with large numbers of parameters and multimodal posterior distributions.

Nested sampling is based on breaking the posterior into a number of nested slices with increasing likelihoods (Figure 5) and generating samples for each. The samples are then

recombined with appropriate weights to yield an estimate of the posterior.

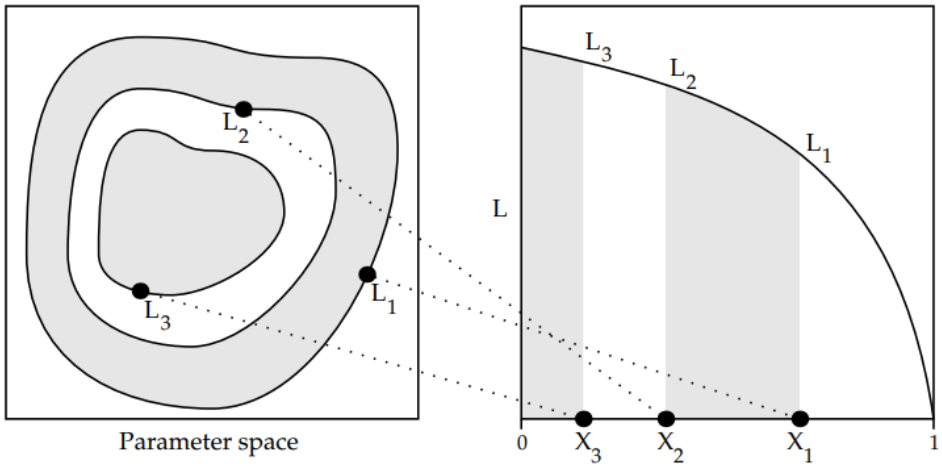


Figure 5: Transformation from parameter space to the one dimensional space of the prior mass (from Skilling (2006)).

Denoting as $X(\lambda)$ the cumulant prior mass corresponding to likelihood values higher than λ and $L(X)$ the inverse of that function, i.e. the likelihood corresponding to the prior mass, the evidence becomes the one dimensional integral given in the equation below.

$$\mathcal{Z} = \int_0^1 L(X)\ \mathrm{d}X \tag{16}$$

Modeling the prior volume corresponding to each dead point as $X_i = t_i X_{i-1}$, where t_i is the shrinkage ratio taken as the largest among n random variables from the interval $[0, 1]$ we get the following expressions for the probability, expected value and variance, respectively:

$$P(t_i) = n \cdot t_i^{n-1} \tag{17}$$

$$E[\ln(t_i)] = -\frac{1}{n} \tag{18}$$

$$\mathrm{Var}[\ln(t_i)] = \frac{1}{n} \tag{19}$$

Samples from the posterior distribution are also obtained simultaneously with the evidence calculation. Initially, a number of live points is randomly sampled from the prior. At each iteration, the point with the lowest likelihood is replaced with a point from a region with higher likelihood and added to a list of dead points. The process is ended when a stopping criterion is satisfied, and afterwards the dead points as well as the current live points are used to estimate the evidence and posterior distribution. The behaviour of the likelihood and prior mass for progressive iterations of the algorithm are shown in Figure 6.

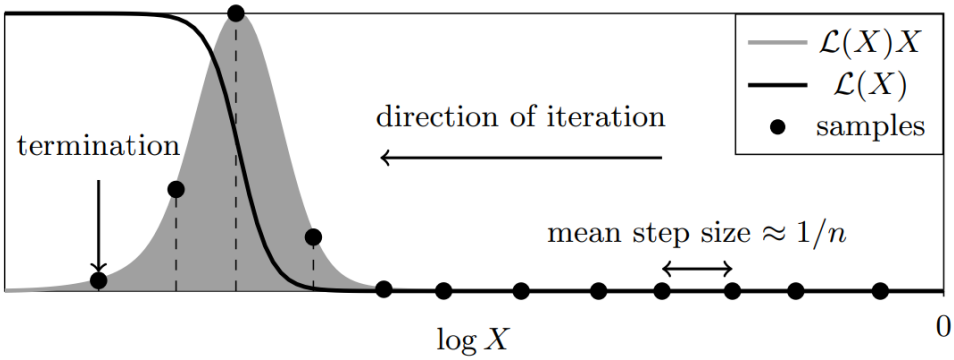


Figure 6: Behaviour of the likelihood and prior mass for progressive sampling iterations (from Higson et al. (2018)).

Denoting the set of shrinkage ratios of the dead points as $\boldsymbol{t} = \{t_1, t_2, ...t_{\mathrm{n_{dead}}}\}$, the evidence is obtained as the quadrature sum over the dead points:

$$\mathcal{Z}(\boldsymbol{t}) \approx \sum_{i \in dead} L_i w_i(\boldsymbol{t}) \tag{20}$$

where the weights w_i are calculated from the trapezium rule:

$$w_i = \frac{1}{2}(X_{i-1}(t) - X_{i+1}(t)) \tag{21}$$

Samples from the posterior can then be obtained using the calculated likelihood L_i and quadrature weights w_i as:

$$p_i(\boldsymbol{t}) = \frac{w_i(\boldsymbol{t})L_i}{\mathcal{Z}(\boldsymbol{t})} \tag{22}$$

The estimation of the evidence at each iteration allows for a meaningful stopping criterion to be set for the algorithm based on an estimate of the remaining evidence. For iteration i, an upper bound on the remaining evidence $\Delta\hat{\mathcal{Z}}_i$ can be constructed as $\Delta\hat{\mathcal{Z}}_i \leq L_{\max} X_i$. The stopping criterion is then formulated as:

$$\ln(\hat{\mathcal{Z}} + \Delta\hat{\mathcal{Z}}_i) - \ln(\hat{\mathcal{Z}}_i) < \epsilon, \tag{23}$$

with ϵ being the tolerance on the upper bound of the remaining evidence, taken as $10^{-3}(K - 1) + 10^{-2}$ where K is the number of live points.

The standard (i.e. static) nested sampling procedure discussed previously has been further extended by Higson et al. (2018) to use dynamically varying number of points, allowing for control over the allocation of computational resources for either evidence calculation or calculation of the posterior by adjusting the sampling parameters. In this thesis the static variant is used, implemented in the *Dynesty* package (Speagle, 2019).

3.5 Information theory

An information theory based approach is adopted to tackle the problem of sensor layout optimization. The information content of different sensor layouts (as well as different sets of measurements) can be estimated by evaluating the information entropy between the prior and posterior distributions. For this purpose the Kullback-Leibler (KL) divergence can be used. The KL divergence, also referred to as the relative entropy (MacKay (2003)), is a measure of the difference between two distributions $p(\theta)$ and $q(\theta)$ and can be obtained as:

$$D_{\mathrm{KL}}(P||Q) = \int_{-\infty}^{\infty} p(\theta) \ln(\frac{p(\theta)}{q(\theta)}) \mathrm{d}\theta \tag{24}$$

Using this definition of the relative entropy, the information gain from the prior $p(\boldsymbol{\theta})$ to the posterior $p(\boldsymbol{\theta}|\boldsymbol{y})$ for a given sensor layout or set of measurements is computed as shown in Equation 25, with the largest entropy corresponding to the optimal layout or the most informative set of measurements.

$$H = \int_{\Theta} p(\boldsymbol{\theta}|\boldsymbol{y}) \ln \frac{p(\boldsymbol{\theta}|\boldsymbol{y})}{p(\boldsymbol{\theta})} \mathrm{d}\boldsymbol{\theta} \tag{25}$$

A full Bayesian parameter estimation is required to obtain the relative entropy for each evaluated layout. This results in a high computational cost when this method is used for sensor placement and measurement selection, either using global optimization or forward sequential sensor placement. To remedy this, an approach based on the differential entropy is applied instead.

3.5.1 FIM approximation

In order to reduce the computational cost, an approximate method based on the Fisher Information Matrix (FIM) is applied, as described in Papadimitriou and Lombaert (2012). In the following, the number of measurements for each observed degree of freedom and the number of observed degrees of freedom are denoted as N and N_0 respectively. The total number of degrees of freedom is denoted as N_d. The following definition of the information entropy is used, referred to as the differential entropy:

$$h(\boldsymbol{L}; \boldsymbol{\Sigma}, \boldsymbol{y}) = E_{\theta}[-\ln p(\boldsymbol{\theta}; \boldsymbol{\Sigma}, \boldsymbol{y})] = -\int_{\boldsymbol{\theta}} \ln p(\boldsymbol{\theta}|\boldsymbol{\Sigma}, \boldsymbol{y}) p(\boldsymbol{\theta}|\boldsymbol{\Sigma}, \boldsymbol{y}) \mathrm{d}\boldsymbol{\theta}, \tag{26}$$

where $\boldsymbol{L} \in R^{N_0 \times N_d}$ is the observation matrix (or sensor configuration matrix) that defines the locations of sensors in the structure. For each row i, a single element j is set to one, representing sensor i placed at the degree of freedom j. All other elements are equal to zero.

The differential entropy provides a measure of the uncertainty in the posterior parameter distributions. Minimizing the information entropy corresponds to minimizing the posterior uncertainty. A drawback of the differential entropy formulation is that it quantifies the information content of the posterior and therefore is dependent on the prior distribution, as shown in Equation 26. This is not the case with the KL-divergence which measures the

information gain from prior to posterior and therefore does not depend on the prior.

For $NN_0 \to \infty$, i.e. when a large number of measurements is available, the relative entropy given in Equation 26 can be approximated by a Laplace type integral (Papadimitriou et al., 2000). The following approximation for the information can be applied:

$$h(\boldsymbol{L}; \boldsymbol{\Sigma}_0, \boldsymbol{y}) \sim H(\boldsymbol{L}; \boldsymbol{\theta}_0, \boldsymbol{\Sigma}_0) = \frac{1}{2} N_\theta \ln 2\pi - \frac{1}{2} \ln[\det Q(\boldsymbol{L}; \boldsymbol{\theta}_0; \boldsymbol{\Sigma}_0)], \tag{27}$$

where

$$Q(\boldsymbol{L}; \boldsymbol{\theta}, \boldsymbol{\Sigma}_0) = \sum_{k=1}^{N} (\boldsymbol{L} \nabla_\theta \boldsymbol{y}_k)^T (\boldsymbol{L} \boldsymbol{\Sigma}_0 \boldsymbol{L}^T)^{-1} (\boldsymbol{L} \nabla_{\boldsymbol{\theta}} \boldsymbol{y}_k) \tag{28}$$

The term $Q(\boldsymbol{L}; \boldsymbol{\theta}_0, \boldsymbol{\Sigma})$ in the previous expression denotes the FIM and $\nabla_{\boldsymbol{\theta}}$ denotes the gradient vector $[\partial/\partial\theta_1, \partial/\partial\theta_2, ..., \partial/\partial\theta_{N_\theta}]$. The vector of observations $\boldsymbol{y}$ is not available when performing sensor optimization and measurement selection. In order to obtain useful designs it is assumed that $\boldsymbol{\theta}_0$ is a vector of nominal parameter values based on prior knowledge or engineering judgement. The vector $\boldsymbol{y}_k$ is taken as the model output for the assumed nominal parameters.

3.5.2 Forward sequential sensor placement

The Forward Sequential Sensor Placement (FSSP) method is applied, as described in Papadimitriou and Lombaert (2012). The structure with no sensors is considered initially. The objective of this iterative procedure is to obtain the vector:

$$\boldsymbol{L}_{\text{opt},i} = \text{argmin}_{\boldsymbol{L}_i} H(\boldsymbol{L}; \boldsymbol{\theta}_0, \boldsymbol{\Sigma}_0), \tag{29}$$

i.e. the observation matrix corresponding to the sensor placement that minimizes the relative entropy at each step i. The first iteration is performed assuming that no sensors are placed on the structure, and the first sensor is placed at the DOF that minimizes the negative information entropy over the set of N_p measurable DOFs. The process is then repeated with an additional sensor being placed at each iteration step. Denoting the number of sensors at step i and the maximum number of sensors to be placed as N_i and N_max respectively, the algorithm for applying FSSP is illustrated below.

Algorithm 1: Forward Sequential Sensor Placement (FSSP)

Result: Observation matrix $\boldsymbol{L}_i$ of optimal sensor positions for step i
N_{init} = Initial number of sensors placed on the structure;
$\boldsymbol{L}_{\text{init}}$ = Initial observation matrix;
Set current number of sensors $N_i = N_{\text{init}}$;
while $N_i < N_{max}$ **do**
  N_{pi} = *Number of unobserved measurable DOFs at iteration i;*
  for $j \leftarrow 0$ **to** N_{pi} **do**
    Define $\boldsymbol{L}_{i,j}$ *containing unobserved DOF j ;*
    Calculate $H_{i,j}$ *corresponding to observation matrix* $\boldsymbol{L}_{i,j}$*;*
    Find $\boldsymbol{L}_{opt,i} = argmin_{\boldsymbol{L}_{i,j}} H(\boldsymbol{L}_{i,j}; \boldsymbol{\theta}_0, \boldsymbol{\Sigma})$ *;*
    Set $\boldsymbol{L}_i = \boldsymbol{L}_{opt,i}$ *;*
  end
  Increase N_i *by one;*
end

3.6 Concluding summary

The methods and tools used to develop examples and perform analyses are summarized. Using the Bayes theorem, the Bayesian system identification framework allows for a mathematically rigorous combination of physical models and measurements to obtain the posterior distribution of parameters. The likelihood function is formulated based on the data generating process, i.e. the coupled physical-probabilistic model describing the process that is assumed to generate the measurements. Different kernel functions are described to model the assumed correlations in the physical model prediction error. A nested sampling approach is then presented, it's utility is to estimate the evidence and posterior. Finally, different measures of the information content of a set of measurements are presented, such as the relative entropy or KL-divergence and the differential entropy, as well as a method to efficiently approximate the former requiring only the evaluation of the model gradient for assumed nominal parameters.

4 Efficient likelihood evaluation

The log-likelihood function of the multivariate normal distribution, assuming zero mean, is given in Equation 30. Evaluation of the loglikelihood requires calculating the inverse and determinant of the covariance matrix $\boldsymbol{\Sigma}$. These operations typically have a complexity of $O(N^3)$ and $O(N^2)$ storage requirements for the covariance matrix, where N is the number of dimensions of the multivariate normal distribution.

$$\log L(\boldsymbol{y}) = -\frac{1}{2}[\log|\boldsymbol{\Sigma}| + \boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y} + N\log 2\pi] \tag{30}$$

In order to reduce the computational complexity of evaluating the likelihood, we propose an approach that utilizes the tridiagonal inverse form of the correlation matrix that can be obtained from the Matern family of kernel functions, as well as the Kronecker structure of the combined space and time covariance matrix. The proposed method is applicable for evaluating the loglikelihood under the multiplicative modeling uncertainty model with additive i.i.d. Gaussian noise described in Section 3.2.2. For the additive uncertainty model described in Section 3.2.1, an approach based on the efficient eigendecomposition of the covariance matrix is implemented.

4.1 Mathematical tools

A brief description of the Kronecker product and tridiagonal matrices, the mathematical tools used throughout this section, is provided below.

4.1.1 Kronecker product

The Kronecker product of two matrices is given as:

$$\boldsymbol{A}\otimes\boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & \dots & a_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & \dots & a_{mn}\boldsymbol{B} \end{bmatrix} \tag{31}$$

Some of the properties of the Kronecker product of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are listed below.

1. The inverse can be obtained as $(\boldsymbol{A}\otimes\boldsymbol{B})^{-1} = \boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1}$.
2. The determinant can be obtained as $|\boldsymbol{A}\otimes\boldsymbol{B}| = |\boldsymbol{A}|^m|\boldsymbol{B}|^n$.
3. Given the eigenvalues $\lambda_1, \lambda_2, ...\lambda_n$ and $\mu_1, \mu_2, ...\mu_m$ for $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively, the eigenvalues of $\boldsymbol{A}\otimes\boldsymbol{B}$ are $\lambda_i\cdot\mu_j$ for $i\in[n], j\in[m]$.
4. Vectorization can be applied as $(\boldsymbol{B}^T\otimes\boldsymbol{A})\text{vec}(\boldsymbol{X}) = \text{vec}(\boldsymbol{AXB})$

4.1.2 Tridiagonal matrices

A tridiagonal matrix is defined as a matrix with non-zero elements only in the main diagonal, as well as the diagonals above and below it:

$$\boldsymbol{C} = \begin{bmatrix} b_1 & c_1 & & & \\ a_1 & b_2 & c_2 & & \\ & a_2 & b_3 & c_3 & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \tag{32}$$

Both the general tridiagonal and the symmetric tridiagonal form offer significant computational advantages due to the fact that several matrix operations can be performed more efficiently, e.g. inversion (Meurant, 1992), eigendecomposition (Coakley and Rokhlin, 2013) and Cholesky decomposition (Bar-On et al., 2006).

4.2 Proposal for efficient likelihood evaluation for multiplicative modeling uncertainty

In the following, exponential correlation is assumed in both space and time. The ij^{th} element of the spatial (or temporal) covariance matrix $\boldsymbol{C}$ is obtained as:

$$\boldsymbol{C}_{ij}(x_i, x_j) = \sigma_{ij}^2 \exp\left(\frac{\|x_i - x_j\|}{l}\right), \tag{33}$$

where l is the correlation length. It is shown in Pasquier and Marcotte (2020) that the inverse of the covariance matrix for this kernel function has a symmetric tridiagonal form:

$$\boldsymbol{C}^{-1} = \begin{bmatrix} d_1 & c_1 & & & \\ c_1 & d_2 & c_2 & & \\ & c_2 & d_3 & c_3 & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \tag{34}$$

This is due to the Markov property of the exponential kernel also holds for other kernel functions (Marcotte and Allard, 2018). Following Cheong (2016), the diagonal vectors of diagonal and off-diagonal terms in Equation 34 can be obtained analytically, eliminating the need to form the full covariance matrices which is often computationally intensive due to the amount of memory and operations required. For a given vector of observations with coordinates $\boldsymbol{x} = \{x_1, x_2, ..., x_n\}$ denoting $\Delta x_i = |x_i - x_{i-1}|$ for $i \in [n]$ yields Equation 35.

$$a_i = e^{-\lambda \cdot \Delta x_i} \tag{35}$$

where:

- λ is the inverse of the correlation length l.
- a_i is the correlation between points i and $i-1$

Letting $\overline{a}_{ij}$ denote the ij^{th} element of the inverse correlation matrix $\boldsymbol{A}(\lambda)^{-1}$ corresponding to $\boldsymbol{x}$, the elements of $\boldsymbol{A}(\lambda)^{-1}$ can be obtained as:

$$\overline{a}_{ii} = \frac{1}{1-a_2^2}, \tag{36}$$

$$\overline{a}_{nn} = \frac{1}{1-a_n^2}, \tag{37}$$

$$\overline{a}_{jj} = \frac{1}{1-a_i^2} + \frac{1}{1-a_{i+1}^2} - 1, \tag{38}$$

$$\overline{a}_{ii-1} = \frac{-a_i}{1-a_i^2}, \tag{39}$$

$$\overline{a}_{ij} = 0, \text{ for } |i-j| > 1. \tag{40}$$

These terms form the diagonal and off-diagonal vectors of the inverse correlation matrix $\boldsymbol{A}(\lambda)^{-1}$ and can be used to directly obtain the corresponding vectors of the covariance matrix. Computationally this reduces memory requirements by storing the matrix as two vectors of lengths N and $N-1$. Furthermore, we consider a combined space and time covariance which can be obtained as the Kronecker product of the space covariance $\boldsymbol{C}_x(\boldsymbol{x}, \boldsymbol{x}')$ and the time covariance $\boldsymbol{C}_t(\boldsymbol{t}, \boldsymbol{t}')$:

$$\boldsymbol{C}((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{t}, \boldsymbol{t}')) = \boldsymbol{C}_x(\boldsymbol{x}, \boldsymbol{x}') \otimes \boldsymbol{C}_t(\boldsymbol{t}, \boldsymbol{t}') \tag{41}$$

Using the properties of the Kronecker product, it can be shown that the resulting inverse matrix $C((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{t}, \boldsymbol{t}'))^{-1}$ has a symmetric block tridiagonal form, where each block is a symmetric tridiagonal matrix:

$$\boldsymbol{C}((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{t}, \boldsymbol{t}'))^{-1} = \begin{bmatrix} \boldsymbol{D}_1 & \boldsymbol{C}_1 & & & \\ \boldsymbol{C}_1 & \boldsymbol{D}_2 & \boldsymbol{C}_2 & & \\ & \boldsymbol{C}_2 & \boldsymbol{D}_3 & \boldsymbol{C}_3 & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \tag{42}$$

It is noted that the tridiagonality and block tridiagonality shown previously still hold under the assumption of multiplicative modeling uncertainty. We consider $\boldsymbol{K} = \boldsymbol{Y} \cdot \boldsymbol{C} \cdot \boldsymbol{Y}^T$ where $\boldsymbol{C}$ has either a tridiagonal or block tridiagonal inverse and $\boldsymbol{Y}$ is obtained as the diagonal matrix of the vectorized matrix of model output $\boldsymbol{Y} = \text{diag}\left(\text{vec}\left(\boldsymbol{Y}_{model}\right)\right)$, where $\boldsymbol{Y}_{\text{model}}$ is the $N_t \times N_x$ matrix of the model prediction. Then $\boldsymbol{K}^{-1}$ will also be tridiagonal or block tridiagonal respectively. However, including additive noise such that $\boldsymbol{K} = \boldsymbol{Y} \cdot \boldsymbol{C} \cdot \boldsymbol{Y}^T + \sigma^2 \cdot \boldsymbol{I}$ leads to a dense inverse.

We consider the covariance matrix for the multiplicative modeling uncertainty model described in Section 3.2.2, expressed as:

$$\boldsymbol{\Sigma} = \boldsymbol{Y}\boldsymbol{C}\boldsymbol{Y}^T + \boldsymbol{W}, \tag{43}$$

where

- $\boldsymbol{\Sigma}$ is the covariance matrix with multiplicative uncertainty plus i.i.d. noise
- $\boldsymbol{Y}$ is the diagonal matrix of model output
- $\boldsymbol{C}$ is the correlation matrix
- $\boldsymbol{W}$ is the diagonal matrix of i.i.d. noise

To efficiently evaluate the likelihood we aim to calculate the terms $\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$ and $|\boldsymbol{\Sigma}|$ without explicitly forming the corresponding matrices or directly inverting the covariance matrix, while taking advantage of the properties described previously to reduce the complexity.

Algebraic manipulation of the product $\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$ is performed in order to obtain an expression that can be evaluated efficiently by taking advantage of the Kronecker structure and block symmetric tridiagonal inverse of the covariance matrix. We apply the Woodbury identity given below:

$$(\boldsymbol{A}^{-1} + \boldsymbol{B}\boldsymbol{C}^{-1}\boldsymbol{B}^T)^{-1} = \boldsymbol{A} - \boldsymbol{A}\boldsymbol{B}(\boldsymbol{C} + \boldsymbol{B}^T\boldsymbol{A}\boldsymbol{B})^{-1}(\boldsymbol{A}\boldsymbol{B})^T \tag{44}$$

Substituting $\boldsymbol{A} \to \boldsymbol{W}^{-1}$, $\boldsymbol{B} \to \boldsymbol{Y}$ and $\boldsymbol{C}^{-1} \to \boldsymbol{C}$, yields:

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{W}^{-1} - (\boldsymbol{W}^{-1}\boldsymbol{Y})(\boldsymbol{C}^{-1} + \boldsymbol{Y}^T\boldsymbol{W}^{-1}\boldsymbol{Y})^{-1}(\boldsymbol{W}^{-1}\boldsymbol{Y})^T, \tag{45}$$

Applying the left and right vector multiplication by $\boldsymbol{y}$, the second term in the r.h.s. of Equation 30 becomes:

$$\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y} = \boldsymbol{y}^T\boldsymbol{W}^{-1}\boldsymbol{y} - \boldsymbol{y}^T(\boldsymbol{W}^{-1}\boldsymbol{Y})(\boldsymbol{C}^{-1} + \boldsymbol{Y}^T\boldsymbol{W}^{-1}\boldsymbol{Y})^{-1}(\boldsymbol{W}^{-1}\boldsymbol{Y})^T\boldsymbol{y} \tag{46}$$

In the previous expression, the term $\boldsymbol{y}^T\boldsymbol{W}^{-1}\boldsymbol{y}$ can be efficiently evaluated as the product of vectors and diagonal matrices. Similarly, the term $\boldsymbol{y}^T(\boldsymbol{W}^{-1}\boldsymbol{Y})$ can be directly computed and yields a vector. We consider the following term from the r.h.s. of Equation 46:

$$(\boldsymbol{C}^{-1} + \boldsymbol{Y}^T\boldsymbol{W}^{-1}\boldsymbol{Y})^{-1}(\boldsymbol{W}^{-1}\boldsymbol{Y})^T\boldsymbol{y} = \boldsymbol{X} \;\Rightarrow\; (\boldsymbol{C}^{-1} + \boldsymbol{Y}^T\boldsymbol{W}^{-1}\boldsymbol{Y})\boldsymbol{X} = (\boldsymbol{W}^{-1}\boldsymbol{Y})^T\boldsymbol{y} \tag{47}$$

For the case of combined spatial and temporal correlation, we note that the term $\boldsymbol{C}^{-1} + \boldsymbol{Y}^T\boldsymbol{W}^{-1}\boldsymbol{Y}$ is the sum of a symmetric block tridiagonal matrix $\boldsymbol{C}^{-1}$ and the diagonal matrix $\boldsymbol{Y}^T\boldsymbol{W}^{-1}\boldsymbol{Y}$. We can therefore take advantage of efficient algorithms for Cholesky factorization of symmetric block tridiagonal matrices and for solving linear systems using this factorization to compute $\boldsymbol{X}$. The factorization yields:

$$\boldsymbol{C}^{-1} + \boldsymbol{Y}^T\boldsymbol{W}^{-1}\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{L}_1 & & & & \\ \boldsymbol{E}_1 & \boldsymbol{L}_2 & & & \\ & \boldsymbol{E}_2 & \boldsymbol{L}_3 & & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{L}_1 & \boldsymbol{E}_1^T & & & \\ & \boldsymbol{L}_2 & \boldsymbol{E}_2^T & & \\ & & \boldsymbol{L}_3 & \boldsymbol{E}_3^T & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \tag{48}$$

We then obtain $\boldsymbol{X}$ by solving the linear system:

$$\begin{bmatrix} \boldsymbol{L}_1 & & & & \\ \boldsymbol{E}_1 & \boldsymbol{L}_2 & & & \\ & \boldsymbol{E}_2 & \boldsymbol{L}_3 & & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{L}_1 & \boldsymbol{E}_1^T & & & \\ & \boldsymbol{L}_2 & \boldsymbol{E}_2^T & & \\ & & \boldsymbol{L}_3 & \boldsymbol{E}_3^T & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \cdot \boldsymbol{X} = (\boldsymbol{W}^{-1}\boldsymbol{Y})^T \boldsymbol{y} \tag{49}$$

The efficiency of this calculation could potentially be further improved by considering the tridiagonality of each individual block in the block matrix $\boldsymbol{C}^{-1} + \boldsymbol{Y}^T\boldsymbol{W}^{-1}\boldsymbol{Y}$.

The Cholesky decomposition of the symmetric block tridiagonal covariance matrix obtained previously is also used to reduce the computational cost of evaluating the determinant $|\boldsymbol{\Sigma}| = |\boldsymbol{W} + \boldsymbol{Y}\boldsymbol{C}\boldsymbol{Y}^T|$. Applying the determinant lemma yields Equation 50.

$$|\boldsymbol{W} + \boldsymbol{Y}\boldsymbol{C}\boldsymbol{Y}^T| = |\boldsymbol{C}^{-1} + \boldsymbol{Y}^T\boldsymbol{W}^{-1}\boldsymbol{Y}| \cdot |\boldsymbol{C}| \cdot |\boldsymbol{W}| \tag{50}$$

The determinant $|\boldsymbol{C}|$ can be calculated efficiently using Property 2 of the Kronecker product given that $|\boldsymbol{C}| = |\boldsymbol{C}_x \otimes \boldsymbol{C}_t|$. Furthermore $|\boldsymbol{W}|$ is a constant diagonal matrix meaning that the determinant can be trivially obtained. Finally we have previously calculated the Cholesky factorization of the term $\boldsymbol{C}^{-1} + \boldsymbol{Y}^T\boldsymbol{W}^{-1}\boldsymbol{Y}$. Using the fact that the determinant of a block triangular matrix is the product of the determinants of its diagonal blocks and the properties of the determinant:

$$|\boldsymbol{C}^{-1} + \boldsymbol{Y}^T\boldsymbol{W}^{-1}\boldsymbol{Y}| = |\boldsymbol{L}\boldsymbol{L}^T| = |\boldsymbol{L}| \cdot |\boldsymbol{L}^T| = |\boldsymbol{L}|^2, \tag{51}$$

and since each block $\boldsymbol{L}_{ii}$ is also triangular, the evaluation of the determinant has been reduced to evaluation of the determinant of each triangular block $\boldsymbol{L}_{ii}$, which is equal to the product of its diagonal elements.

Using the procedure described previously, an efficient solution can also be obtained for the case of only spatial or temporal correlation. $\boldsymbol{C}^{-1}$ has the symmetric tridiagonal form given in Equation 34. The term $\boldsymbol{C}^{-1}+\boldsymbol{Y}^T\boldsymbol{W}^{-1}\boldsymbol{Y}$ is the sum of a symmetric tridiagonal and a diagonal matrix. Computationally, this property is advantageous as it allows for the solution to the system of equations (Equation 47) to be obtained with $O(N)$ operations using the Thomas algorithm (Quarteroni et al., 2007). Alternatively, for improved efficiency and numerical stability a Cholesky decomposition can be applied to solve the linear system and calculate the determinants of the symmetric tridiagonal terms in Equation 50.

4.2.1 Minimal example in 1D

Timing comparisons are performed to determine the performance gain when using the efficient log-likelihood calculation methods detailed in Section 4. The *SciPy* package developed by Virtanen et al. (2020) was used as a baseline. The reference *SciPy* solution will be referred to as the 'Naive' solution as it does not account for the sparse inverse or Kronecker structure of the covariance matrix. It is also noted that all solutions were obtained on the same machine with an Intel Xeon Platinum 8272CL CPU at 2.60 GHz and 16 GB of RAM and no parallelization is implemented. Under the assumption of exponential correlation, the

log-likelihood of the different synthetic observations was evaluated a total of 50 times per set and the average wall clock time was recorded. It is important to note that for the naive evaluation, the time needed to formulate the correlation and covariance matrices is also included.

The reference *SciPy* solution is compared with the efficient approach that utilizes the tridiagonal form of $\boldsymbol{C}^{-1}$. The average wall clock time as a function of the number of points is plotted in Figure 7. The log-likelihood is evaluated for up to 10^7 points for the proposed approach and 2500 points for the reference solution.

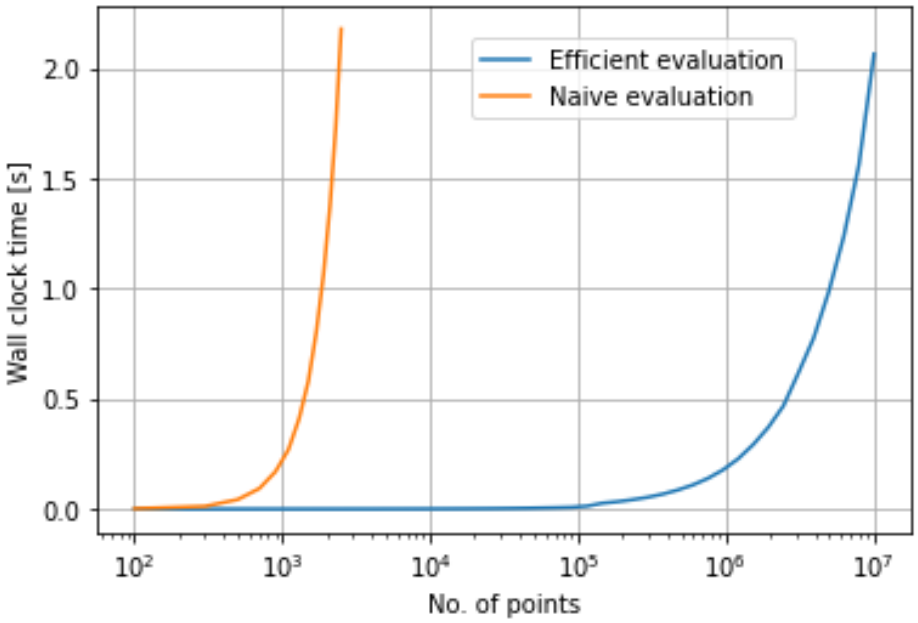


Figure 7: Mean wall clock time comparison of the efficient approach using tridiagonality and naive evaluation.

It is evident that the $O(N)$ complexity achieved by the proposed approach is a significant reduction in computational cost, allowing for exact evaluation of the likelihood in less than one second for datasets of size $N > 10^6$.

4.2.2 Minimal example in 2D

A comparison of the efficient block Cholesky algorithm and the reference *SciPy* solution is performed using the hardware described in the previous example. Grids of two-dimensional synthetic data with exponential correlation were generated with arbitrary uncertainty parameters. The grid sizes range from 8×8 up to 64×64. The results are shown in Figure 8. The color of each pixel represents the average wall clock time for a grid of size $N_x \times N_t$ as denoted by the x and y axis respectively. As with the previous example, the time needed to formulate the correlation and covariance matrices is also included.

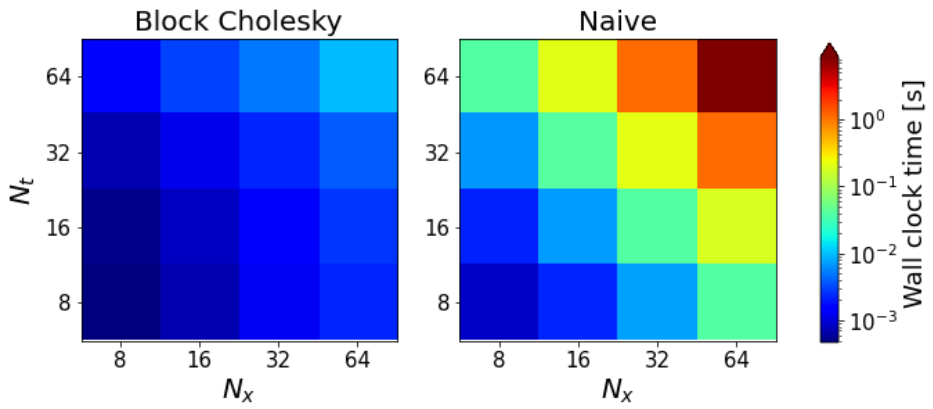


Figure 8: Mean wall clock time comparison of block Cholesky and naive evaluation.

The computational speed-up factor, calculated by dividing the corresponding evaluation times for the different solution methods, for each grid size is illustrated in Figure 9.

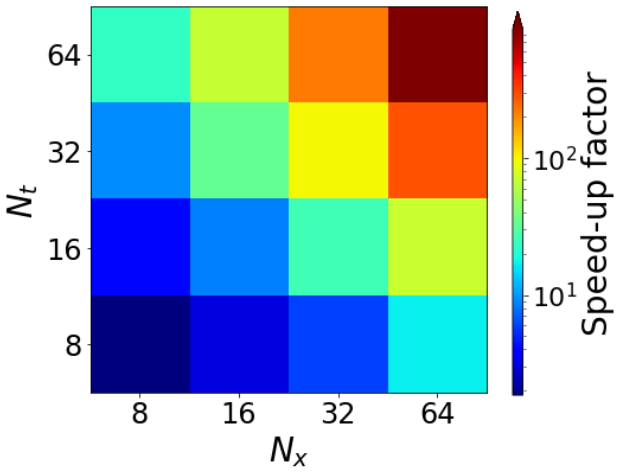


Figure 9: Mean computational speed-up factor for the block Cholesky solution compared to naive evaluation.

The block Cholesky approach results in approximately 900 times faster likelihood evaluation, with an average of 9.8×10^{-3} seconds required for each evaluation on a 64×64 grid. This improvement in efficiency makes it possible to feasibly apply Bayesian inference to significantly larger datasets. Given that, on average, $2 \times 6 \cdot 10^4$ log-likelihood evaluations are needed for a single Bayesian inference in the analyses presented in Section 6, the total log-likelihood evaluation time is reduced from days to minutes using the approaches detailed previously for the case of 9 sensors with 165 measurements each. Additionally, in the block Cholesky approach only the time covariance matrix C_t is formed, which results in reduced memory requirements.

It is important to note that the evaluation time does not scale symmetrically for N_x and N_t, which is also evident by the previous plots. The implemented algorithm performs $3 \cdot N_x$ solves of linear systems of size N_t, as well as N_x Cholesky decompositions of $N_t \times N_t$ sized

matrices. Additionally, the tridiagonal inverse structure of the covariance in the case of the multiplicative modeling uncertainty has not been fully utilized and further efficiency improvements are possible. Furthermore, the aim of the implementation is to showcase the feasibility of the proposed approach and lacks optimization. This can also be achieved by implementing the proposed multiplicative uncertainty approach in a low-level language, e.g. Fortran or C. Details of the implementation can be found in Appendix C.

4.3 Efficient likelihood for additive modeling uncertainty

In the case of additive modeling uncertainty under the assumption of separable space and time covariance, the Kronecker structure of the covariance matrix can be exploited to efficiently evaluate the likelihood $L(\boldsymbol{X}|\boldsymbol{\theta})$ where $\boldsymbol{X} \in^{N_t \cdot N_x}$ denotes the matrix of observations. This type of model is known as a Matrix-variate Gaussian model and has been applied in multi-task Gaussian processes. The full derivation can be found in Stegle et al. (2011). Denoting the space and time covariance matrices as $\boldsymbol{C}_x$ and $\boldsymbol{C}_t$ respectively, the vector property of the Kronecker product can be applied to obtain:

$$(\boldsymbol{C}_x \otimes \boldsymbol{C}_t)\text{vec}(\boldsymbol{X}) = \text{vec}(\boldsymbol{C}_t^T \boldsymbol{X} \boldsymbol{C}_x), \tag{52}$$

where $\text{vec}(\boldsymbol{X})$ denotes the vectorization of $\boldsymbol{X}$. The full covariance matrix for correlated additive modeling uncertainty and i.i.d. Gaussian measurement noise can be written as:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\text{model}} + \boldsymbol{\Sigma}_{\text{meas}} = \boldsymbol{C}_x \otimes \boldsymbol{C}_t + \sigma^2 \boldsymbol{I} \tag{53}$$

Given the eigenvalue decompositions $\boldsymbol{C}_x = \boldsymbol{U}_x \boldsymbol{S}_x \boldsymbol{U}_x^T$ and $\boldsymbol{C}_t = \boldsymbol{U}_t \boldsymbol{S}_t \boldsymbol{U}_t^T$, Equation 53 can be written as:

$$\boldsymbol{\Sigma} = \boldsymbol{C}_x \otimes \boldsymbol{C}_t + \sigma^2 \boldsymbol{I} = (\boldsymbol{U}_x \otimes \boldsymbol{U}_t)(\boldsymbol{S}_x \otimes \boldsymbol{S}_t + \sigma^2 \boldsymbol{I})(\boldsymbol{U}_x^T \otimes \boldsymbol{U}_t^T), \tag{54}$$

Using this decomposition, the log-likelihood can be efficiently obtained as follows:

$$L = -\frac{N_x N_t}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{S}_x \otimes \boldsymbol{S}_t + \sigma^2 \boldsymbol{I}| - \frac{1}{2}\text{vec}(\boldsymbol{U}_t^T x \boldsymbol{U}_x)^T (\boldsymbol{S}_x \otimes \boldsymbol{S}_t + \sigma^2 \boldsymbol{I})^{-1}\text{vec}(\boldsymbol{U}_t^T x \boldsymbol{U}_x) \tag{55}$$

In the case of exponential space and time covariances, the efficiency of this method can be further improved by considering the tridiagonal structure of the inverse covariance matrices and the properties of eigenvalues and eigenvectors. For a matrix $\boldsymbol{A}$ with eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$:

1. The eigenvalues of $\boldsymbol{A}^{-1}$ are $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, ..., \frac{1}{\lambda^n}$
2. The eigenvalues of $\boldsymbol{A} + \alpha \boldsymbol{I}$ are $\lambda_1 + \alpha, \lambda_2 + \alpha, ..., \lambda_n + \alpha$

Combining the properties of eigenvalues and the Kronecker properties of the covariance matrix allows for efficiently evaluating Equation 55 by computing the eigendecompositions of the tridiagonal matrices $\boldsymbol{C}_x^{-1}$ and $\boldsymbol{C}_t^{-1}$.

4.3.1 Minimal example

An analysis of the log-likelihood evaluation time is performed for the case of additive modeling uncertainty, with the method derived by Stegle et al. (2011) and detailed in Section 3.2.1. In the following it will be referred to as the *Kronecker* method for brevity, as it takes advantage of the Kronecker structure of the covariance matrix. A comparison of evaluation wall clock times and speed-up factors is shown in Figure 10 and Figure 11 respectively. Note that the discrepancies in wall clock time for the naive evaluation between the additive and multiplicative (see Section 4.2.2) modeling uncertainty cases are due to the additional matrix multiplications required in the latter. As for the previous cases, the formation of the correlation and covariance matrices is included in the calculation of the wall clock time.

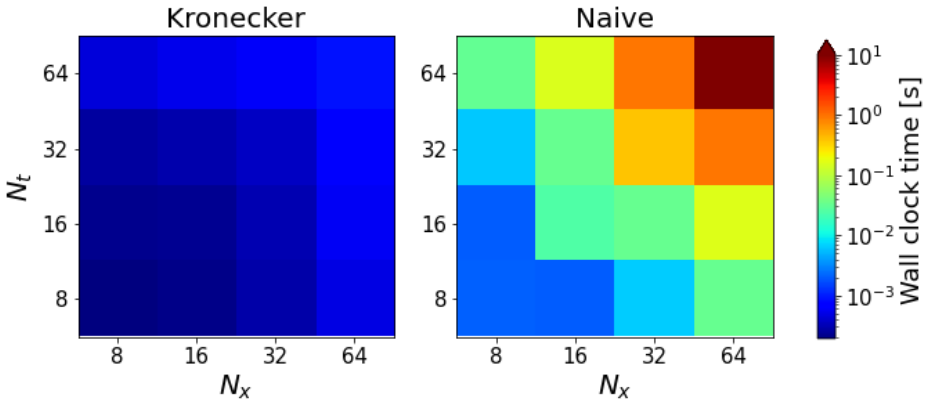


Figure 10: Mean wall clock time comparison of the "Kronecker" approach and naive evaluation.

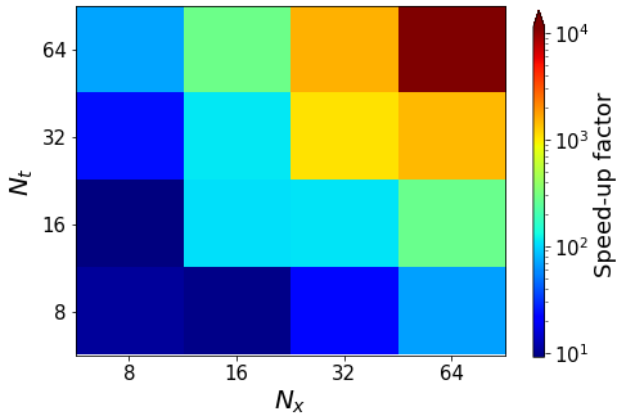


Figure 11: Mean computational speed-up factor for the "Kronecker" solution compared to naive evaluation.

In the Kronecker method neither the space and time covariance matrices C_x and C_t, nor the combined covariance matrix $C_x \otimes C_t$ are formed explicitly. This results in significant reduction of the memory requirements which is desired in large applications. For the 64×64

grid an average evaluation time of $9.1 \cdot 10^{-4}$ s was obtained, while the reference solution required approx. 10.3 s. The calculated speed-up factor is $1.1 \cdot 10^4$. These results are only indicative as they are hardware and implementation dependent. In that respect a more rigorous theoretical analysis of the complexity is needed.

4.4 Concluding summary

An approach is proposed for the efficient evaluation of the Gaussian log-likelihood under spatial and temporal dependencies with multiplicative modeling uncertainty. Taking advantage of the tridiagonal inverse of the covariance matrix for the Matern family of kernels, it is shown that when the combined spatial and temporal covariance matrix can be written as a Kronecker product of the separate spatial and temporal covariance matrices then a block symmetric tridiagonal matrix is obtained, where each block is also symmetric tridiagonal. The inverse and determinant terms in the Gaussian log-likelihood function are reformulated to obtain expressions that can be efficiently evaluated when Gaussian i.i.d. noise is included and subsequently solved using a block tridiagonal Cholesky decomposition. A similar approach allows for exact likelihood evaluation with $O(N)$ complexity for a single (spatial or temporal) dimension. Finally, an eigendecomposition approach is implemented for the case of additive modeling uncertainty in 2D.

The proposed block Cholesky method is shown to outperform the naive evaluation, being approximately 900 times faster on a 64×64 grid, while in the case of additive uncertainty, the eigendecomposition method was found to be more than 10^4 times faster than naive evaluation. In the 1D case a sub-second log-likelihood evaluation time is achieved for more than 10^6 points. The issue of storing large covariance matrices is addressed by utilizing the symmetric tridiagonal inverse form of the exponential covariance matrix to store it as two vectors of length N and $N-1$. These improvements provide a significant reduction in computational cost when exponential correlation is assumed and make inference feasible for more than 10^4 observations.

5 Description of the IJssel bridge case study

5.1 Structure

The IJssel bridge is a steel bridge that carries road traffic over the river IJssel in the direction of Westervoort. It consists of an approach bridge and a main bridge, of which the latter is of interest in this case. It has a total length of 295 m and five spans with lengths of 45, 50, 105, 50 and 45m. In total the bridge has 12 supports. The supports on pillar H are hinges, while the rest are roller bearings in the longitudinal directions. The roller bearings in pillars G and K can resist uplift forces. The main structure of the bridge is composed of two steel girders with variable height, ranging from 2.4 to 5.3 m, and cross-beams with a spacing of approximately 1.8 m. An elevation view of the structure is shown in Figure 12

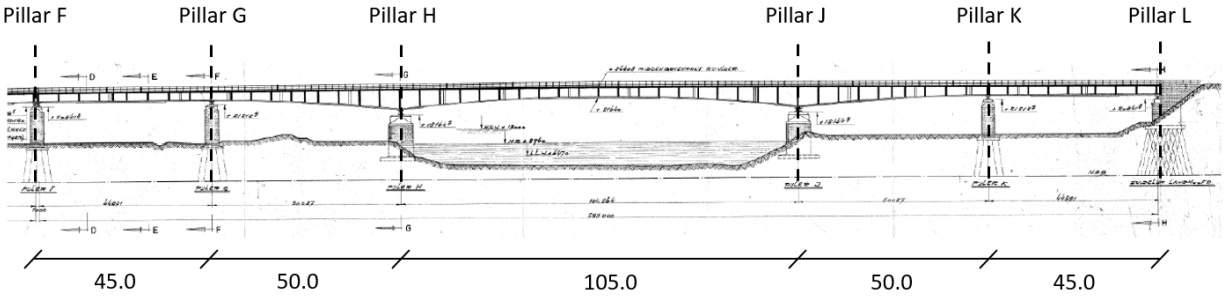


Figure 12: Elevation view of the IJsselbridge.

The main girders and cross-beams support the steel plate deck. The deck has a thickness of 10 or 12 mm and 160 × 8 mm longitudinal bulb stiffeners. The cross beams are placed with a center-to-center distance of 1.75 to 1.80 m and are composed of a 500 × 10 mm web with a 250 × 12 mm welded flange. The cross beams are tapered in the parts that extend beyond the main girders and the beam height is reduced to 200 mm at the beam ends. A cross section view of the bridge is provided in Figure 13.

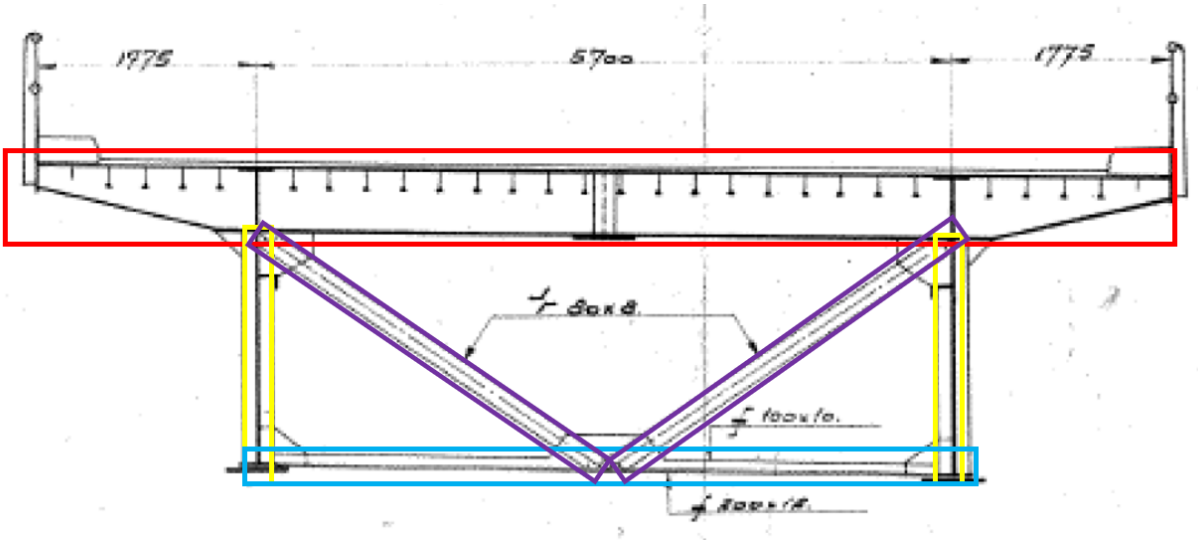


Figure 13: Cross section of the IJsselbridge with highlighted deck (red), K-bracing (purple), bottom cross beam (blue) and main girders (yellow).

The two main girders are coupled with K-braces located below every second or third cross beam. On average the distance between the K-braces is approx. 5.4 m. The K-braces are composed of the elements highlighted in Figure 13:

- Diagonals: Composed of two L sections with dimensions L$80 \times 80 \times 8$ and L $90 \times 90 \times 9$
- Bottom edge: Half DIN 20 profile.
- Verticals: Two half INP24 or INP30 profiles.

5.2 Measurements

The data used in this thesis was obtained from two separate measurement campaigns. During the campaigns several response quantities of the structure were measured including strain, acceleration and temperature. Strain influence lines were obtained for moving loads on the left and right lanes of the bridge using strain gauges at various positions in the structure. The bridge was loaded by trucks of known mass, driven over each of the two lanes at a constant velocity. A subset of the collected data is used in this thesis. The relevant details for each measurement campaign, as well as the data processing are provided below.

5.2.1 TNO measurements

The first set of measurements are obtained from a measurement campaign performed in September 2018. Three strain gauges were placed on the bridge, with two on the center of the bottom flange of the girder closest to the right lane[3] and one at 15 mm from the edge of the flange. The spacing of the sensors at the center of the flange is 200 mm. Measurements from the closely spaced sensors were compared to verify the results. In this thesis only a

[3]Left and right are defined for the direction from pillar F towards pillar L.

subset of the measurements is used, namely the strain recorded by sensor PBR-P4 located at a distance of 65.37 m from the start of the bridge as shown in Figure 14.

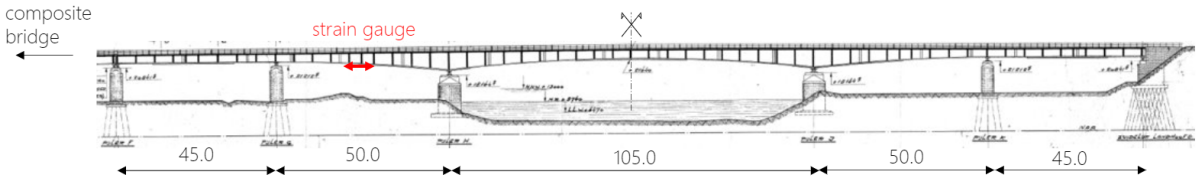


Figure 14: Location of the PBR-P4 strain gauge for TNO measurements (figure taken from internal TNO report).

Trucks with known weights are driven over the bridge at a constant speed of 50.0 km/h in different configurations:

- Single truck on the right lane
- Single truck on the left lane
- One truck per lane simultaneously

The properties of the trucks used for loading the bridge are summarized in Table 3. Additional details on the measurement procedure such as details on the trucks and assumptions regarding the loading conditions can be found in Appendix B

Table 3: Properties of left and right lane trucks used in the TNO measurement campaign.

Truck lane	Axle No.	Axle distance [m]	Load per axle [kN]
Right	1	1.94	57.52
	2	2.09	105.45
	3	1.35	105.45
	4	1.25	105.45
	5	-	105.45
Left	1	2.00	58.86
	2	1.82	107.91
	3	1.82	107.91
	4	1.82	107.91
	5	-	107.91

The stress influence lines for sensor PBR-P4 for truck load on the left and right lanes are shown in Figure 15. The influence line for simultaneous loading of both lanes is not used and therefore omitted. The four manually selected peaks of the influence lines also shown in the figure are used for inference in the reference case discussed in a later section.

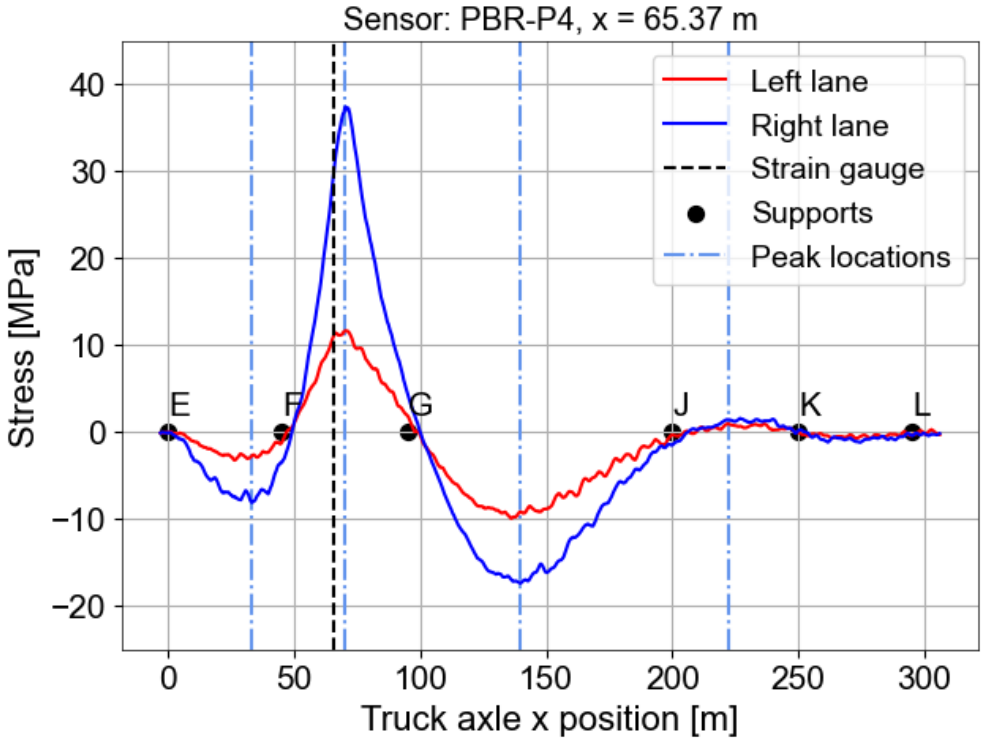


Figure 15: Measured stress influence lines by the TNO sensor PBR-P4.

5.2.2 Fugro measurements

An additional set of measurements from a separate measurement campaign performed by Fugro is used in this thesis. In total, 34 fiber optic strain sensors were placed on the IJsselbridge to measure the response of the structure to traffic load. These sensors measure both strain and temperature on the main steel bridge structure, as well as the concrete-steel composite approach bridge which is not studied in this thesis. The sensors are placed on the top and bottom flange of the steel girders, the cross beams and on the longitudinal bulb stiffeners. The subset of sensors that are placed on the center of the bottom flange of the right main girder is used. These are denoted with the prefix "H". The locations of these sensors is shown in the figure below. Sensors H6 and H11 are located on the left girder and therefore not considered.

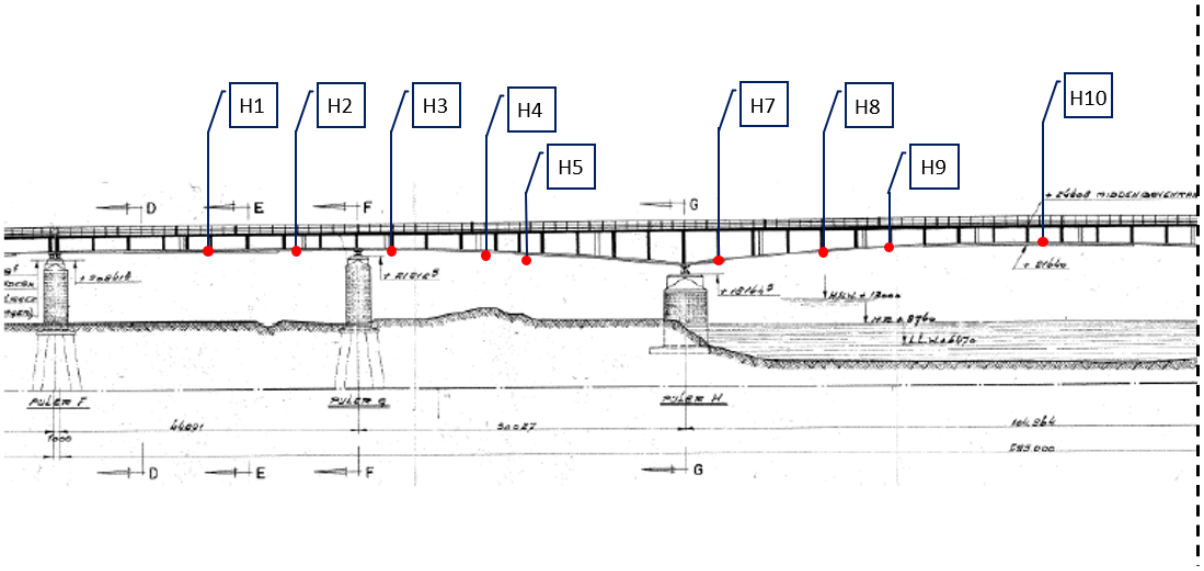


Figure 16: Sensors on the right girder for the Fugro measurement campaign. The prefix "H" is used to denote the sensors on the main structure of the IJsselbridge.

To ease the interpretation of the results in the cases where these sensors are used (see Section 6.4), descriptive labels are assigned to each sensor. The exact position of each sensor along the length of the bridge, as well as the sensor label and new name are provided in Table 4. Only sensors placed in the center of the bottom flange of the main girders are considered and the remaining sensors are not included in the table.

Table 4: Names, labels and positions of Fugro sensors placed on the IJsselbridge main girders.

Sensor	Lane	Position [m]	Label
H1	Right	20.42	H1_20.42_R
H2	Right	34.82	H2_34.82_R
H3	Right	47.70	H3_47.70_R
H4	Right	61.97	H4_61.97_R
H5	Right	68.60	H5_68.60_R
H6	Left	68.60	H6_68.60_L
H7	Right	96.80	H7_96.80_R
H8	Right	113.35	H8_113.35_R
H9	Right	123.90	H9_123.90_R
H10	Right	147.50	H10_147.50_R
H11	Left	147.50	H11_147.50_L

A loaded truck with known properties is driven over the bridge at a constant speed. A single test per lane is performed at a speed of 20 km/h to simulate loading of the bridge under static conditions. Two additional tests per lane at 80 km/h were also performed to measure the structural response under dynamic loading conditions. These are not considered in this

thesis as the scope is limited to the static case. The load per axle and between-axle distances for the trucks are shown in Table 5.

Table 5: Properties of truck used in Fugro measurements.

Axle no.	Axle distance [m]	Load per axle [kN]
1	2.06	59.35
2	1.83	108.82
3	1.82	108.82
4	1.82	108.82
5	-	108.82

Time series of the strain are obtained for each sensor and postprocessed to obtain the corresponding influence lines. These are converted to stress influence lines under the assumption of linear elasticity (Equation 56):

$$\sigma = E \cdot \epsilon, \tag{56}$$

where:

- σ is the stress
- E is the Young's modulus
- ϵ is the strain

The Young's modulus is taken as 210 GPa as specified in the IJsselbridge design. Additional details on the measurement processing procedure, assumptions, and manually selected peaks used in the reference system identification case are provided in Appendix B. The processed influence lines are plotted in the figure below.

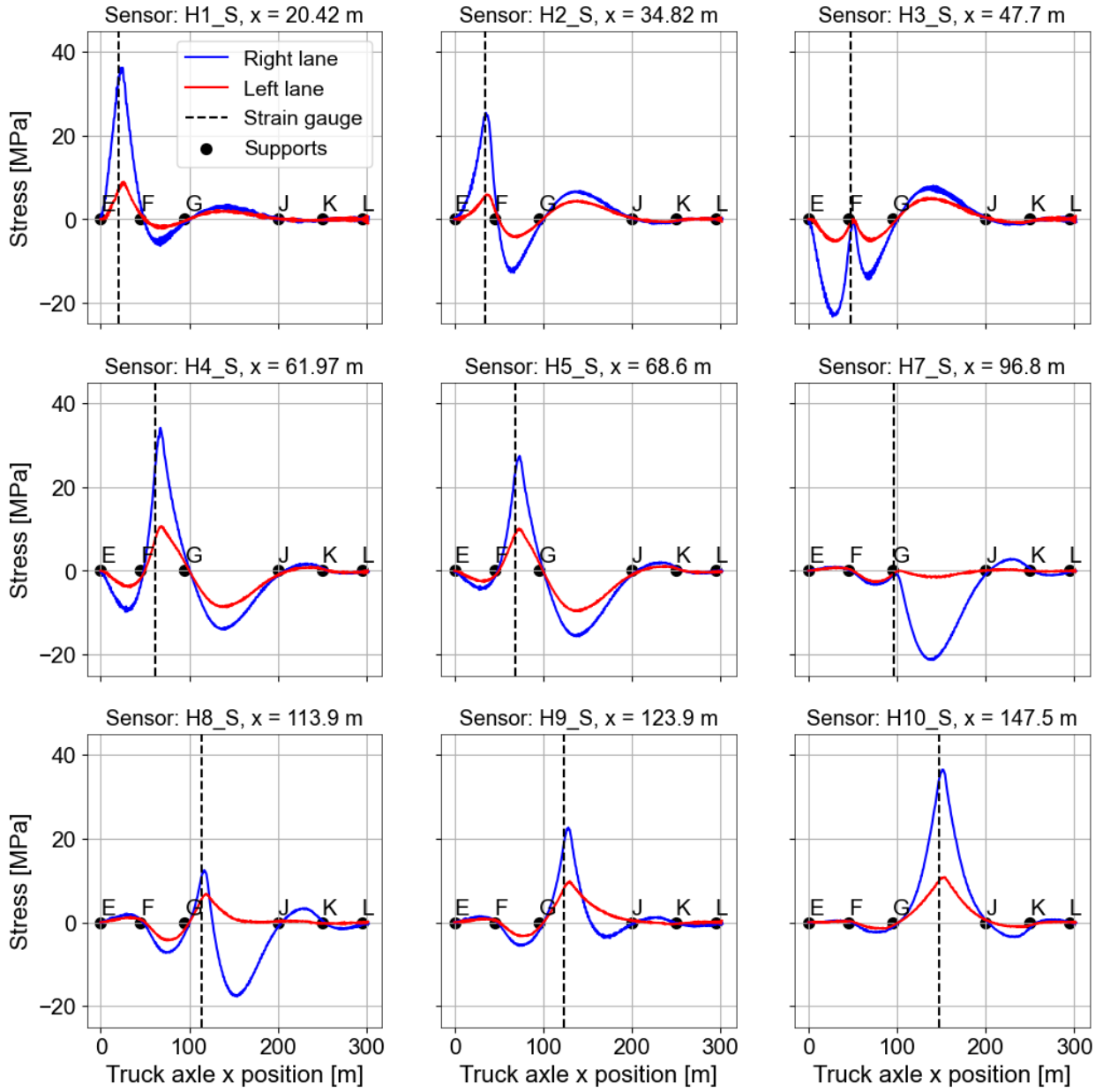


Figure 17: Stress influence lines obtained from the Fugro measurements.

It is important to note that during the processing of the Fugro measurements, large discrepancies between the measurements and the FE model predictions were observed and the structural behaviour in the influence lines could not be replicated by either the single or twin girder FE models described in Section 5.3. For the influence lines from sensors H3 and H7 this can be attributed to the placement of the sensor at a distance less than the beam height from the support, meaning that Euler-Bernoulli beam theory is not applicable. Additionally, qualitative differences in the influence lines between the model and measurements were observed for sensors H8 and H9. These differences could indicate that variations in the truck speed are not correctly accounted for in the measurement processing, that the transfer of forces between the left and right plate girder is not accurately modeled, or that

there is spatial variation of the structural properties in the longitudinal direction that is not accounted for. Parametrizing the twin girder FE model by the full set of input parameters (see Section 5.3.3) and using optimization to fit the model to the full dataset resulted in poor model fit for the aforementioned sensors. The following list is a summary of possible causes for the observed discrepancies:

- Inability of the model to fully capture the structural behaviour;
- Insufficient knowledge of, or incorrect assumptions about the test conditions;
- Variations in the structural behaviour due to the condition of the bridge, e.g. damage to the roller bearings.

The implications of these discrepancies for the analysis are discussed in Section 6.

5.3 Finite element model description

The IJssel bridge is simulated using 2D finite element models based on the Euler-Bernoulli beam element formulation (Bathe, 2006) with four degrees of freedom (DOFs), shown in Figure 18.

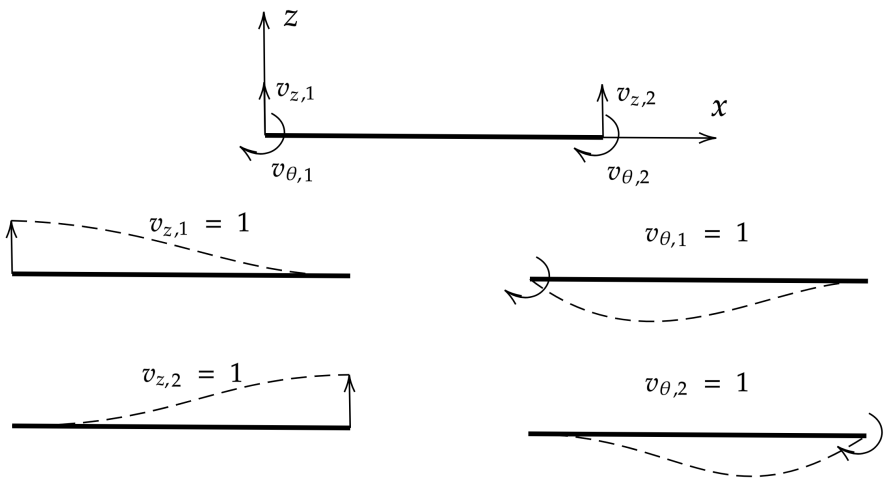


Figure 18: Shape functions of the 4-DOF Euler-Bernoulli beam element.

In this formulation, each node has two degrees of freedom: a rotation and a vertical translation. The structure is discretized into n_{elems} elements and n_{nodes} nodes, with a total of $2 \cdot n_{\text{nodes}}$ DOFs (disregarding the boundary conditions). Obtaining the stress influence line for the truck load moving across the bridge requires calculation of the displacement field for N load time steps[4]. This requires N solutions of the equilibrium equation:

[4]For the static analyses considered in this thesis, the term "time step" is used to refer to each position of the moving load.

$$\boldsymbol{F}_{\mathrm{i}} = \boldsymbol{K} \cdot \boldsymbol{u}_{\mathrm{i}}, \tag{57}$$

where:

- $\boldsymbol{F}_{\mathrm{i}}$ is the nodal load vector at step i;
- $\boldsymbol{K}$ is the global stiffness matrix;
- $\boldsymbol{u}_{\mathrm{i}}$ is the nodal displacement vector at step i;
- i denotes the loading step.

The variable geometrical properties of the steel girders in the longitudinal axis are taken into account by varying the structural properties of the individual beam elements, where each element has a prismatic cross section. The maximum beam element length can be specified in order to approximate the variable geometry to the required precision. Unless otherwise specified, a maximum element length of 2.0 m was used for all simulations. In addition to the main girder, half the width of the deck and the corresponding longitudinal stiffeners are also considered in the calculation of the structural properties for each cross section along the x axis. The resulting model is equivalent to half of the full bridge cross section.

The nodal load vector $\boldsymbol{F}_i$ is obtained by considering the truck loads from individual tires as point loads applied directly to the vertical DOFs corresponding to each tire position. It is evident from Equation 57 that calculating the influence line requires solving N linear systems of $2 \cdot n_{\mathrm{nodes}}$ equations. The Betti-Maxwell theorem (Ghali and Neville, 2017) is utilized to reduce the computational cost to solving a single system of equations of the same size. This allows for the influence line to be obtained with only a single solution of the linear system.

5.3.1 Lateral load function

To account for the position of the truck along the transverse direction (z-axis) which is not included in the FE model, each load is multiplied by the value of the Lateral Load Function (LLF). The LLF is taken as the first order polynomial:

$$f_{\mathrm{LLF}}(x, z) = c_1 \cdot z + c_0 \tag{58}$$

The value of the LLF for a given pair of coordinates (x, z) is the percentage of the load that is transferred to the simulated girder at that position. This is illustrated in Figure 19: A load equal to $f_{\mathrm{LLF}}(x_F, z_F) \times F$ is applied to the simulated girder to take into account the offset position of load F in the transverse direction. For a load applied at the center of the deck ($z = 0$) m, it is assumed that each girder carries half the load. Therefore the slope parameter will be $c_0 = 0.5$ and $f_{\mathrm{LLF}}(x, 0) = 0.5$.

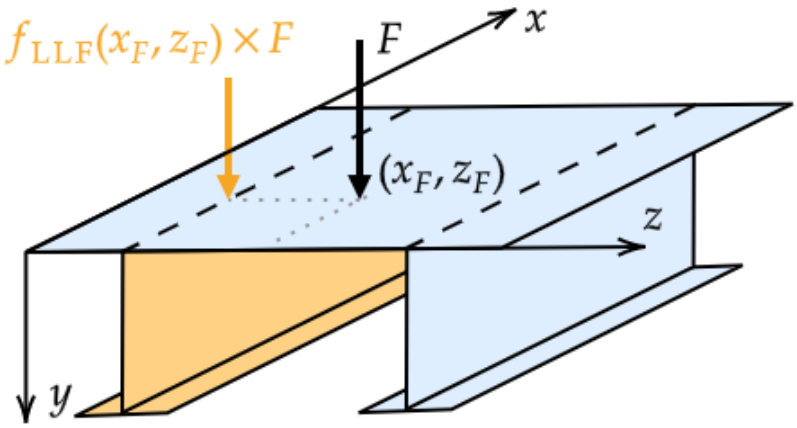


Figure 19: Illustration of the lateral load function.

Although the LLF can be defined such that it accounts for the longitudinal position of the load, this is not applied in this thesis and the dependence on the x coordinate can be omitted from the notation.

5.3.2 Single girder model

The main FE model used is a single girder model with six pinned supports, corresponding to pillars F through L as shown in Figure 12. The assumption of pinned supports is generally a simplification made for modeling and design purposes and is only an approximation of the actual support conditions. In order to account for the possibility of partial locking of the supports, rotational springs are defined at supports F, G, H and J for certain analyses as shown in Figure 20.

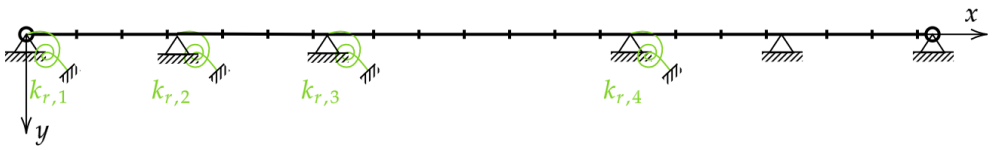


Figure 20: Parametrization of the single girder FE model of the IJsselbridge.

5.3.3 Twin girder model

An additional two-dimensional FE model composed of a system of twin girders, coupled with vertical translation springs is also defined[5]. The same nodal coordinates are used for the nodes of both girders and the transverse direction is not simulated, as shown in Figure 21. The node numbering as well as the vertical translation springs are also shown in the same figure.

[5]It is noted that within this thesis, the twin girder model is only used to investigate the discrepancies observed between the Fugro measurements and model prediction (see Appendix B). The description is included here for completeness

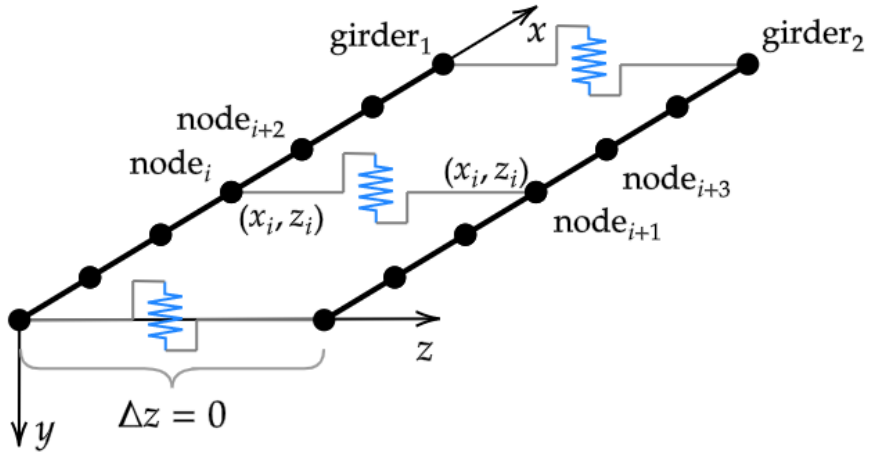


Figure 21: Twin girder FE model of the IJsselbridge.

Supports with rotational springs are defined for the individual girders at locations corresponding to the pillars of the IJssel bridge. The boundary conditions are determined by the stiffness of the rotational springs and can range from pinned, for a spring stiffness of zero, to fully fixed for for very large values of the stiffness. Similarly. the stiffness of the vertical spring determines the degree of coupling between the two beams and can range from complete independence between the two beams to full coupling. The vertical springs are placed approximately at the positions of the K-braces that connect the two main girders of the IJssel bridge. For the analyses performed in this thesis the model is typically parametrized by a single vertical spring stiffness applied to all vertical springs, as well as rotational spring stiffnesses at the second and third support as shown in Figure 22. The twin girder formulation of the model also allows for different rotational spring stiffnesses to be defined for each girder.

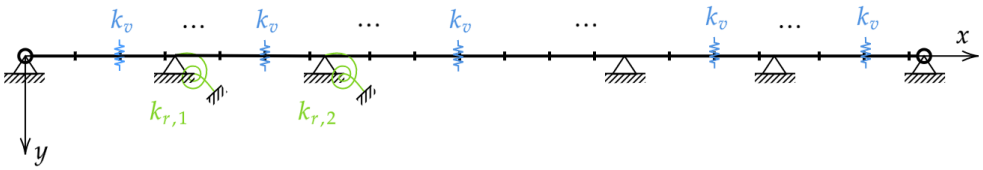


Figure 22: Parametrization of the twin girder FE model of the IJsselbridge.

These model formulations and parametrizations allow for models that can capture more complicated structural behaviour and account for likely damage scenarios, while also offering flexibility and computational efficiency. This allows for developing the entire FE model calculation procedure and integrating it programmatically with the probabilistic analyses presented in the following Chapters. It is thus possible to avoid the need for commercial FEM packages which limit the applicability of parallelization, while also eliminating the additional computational overhead of initializing external FE solvers and performing a large number of I/O operations to disc.

5.3.4 Physical model parameters to be estimated

The set of physical model parameters to be inferred is denoted as $\boldsymbol{\theta}_{\mathrm{s}}$. The physical model parametrization varies between analyses. An overview of the structural parameters is provided in the table below. For the twin girder model, different rotational stiffnesses can be specified for the right and left girder supports. In that case the corresponding rotational stiffnesses are denoted as $K_{\mathrm{ri}}^{\mathrm{r}}$ and $K_{\mathrm{ri}}^{\mathrm{l}}$ respectively. For the single girder model the superscripts are omitted.

Table 6: Description of parameters for the IJsselbridge FE model.

Symbol	Description	Unit
c_1	Lateral load function slope	-
$K_{\mathrm{ri}}^{\mathrm{r}}$	Rotational stiffness of the i^{th} support of the right girder	kNm/rad
$K_{\mathrm{ri}}^{\mathrm{l}}$	Rotational stiffness of the i^{th} support of the left girder	kNm/rad
K_{v}	Vertical stiffness of K braces connecting the right and left girder	kN/m

The prior distributions of the physical parameters are determined using a combination of prior knowledge and engineering judgement. For the slope parameter c_1 of the LLF given in Equation 58, a uniform prior distribution in the range $[-1.0, 1.0]$ is used [6].

For the rotational stiffness of the supports it is suspected that partial locking may be present. It is assumed that any stiffness value between zero (pinned support) and infinity (fixed support) is equally likely. In practice the stiffness must be finite and therefore a very high value is specified instead. This is determined by comparing the stress influence lines for increasing values of the rotational stiffness. The upper bound of the support for the prior distribution of the rotational stiffness is chosen as the point where any increase has a limited effect on the calculated influence line. Details on the sensitivity of the model response to the structural parameters are provided in Appendix A.

For the vertical springs representing the K-braces that connect the two girders, the support of the prior is determined in a similar manner. Unless otherwise specified, the priors given in the table below are used for the physical model parameters.

Table 7: Prior distributions of physical model parameters. Units are given in Table 6.

Parameter	Distribution
c_1	$\mathcal{U}(-1.0, 1.0)$
K_{r}	$\mathcal{U}(0.0, 1.0 \cdot 10^8)$
K_{v}	$\mathcal{U}(0.0, 1.0 \cdot 10^6)$

[6] This support allows for values of the LLF that are not possible in practice. However, it was chosen as a common support for all of the coefficients of a polynomial LLF of order n.

5.4 Probabilistic model description

5.4.1 Covariance formulation

The general formulation of the covariance matrix is described for a Gaussian process in D dimensions in Section 3.3.2, with the position of each observation described by a D-dimensional vector $\boldsymbol{x} = [x_1, x_2, ..., x_D]$. For the case of the IJsselbridge, the coordinates of each observation can be defined by the position of the sensor and the position of the load along the longitudinal axis of the bridge, denoted as x_1 and x_2 respectively.

$$\boldsymbol{x}_i = [x_{i,1}, x_{i,2}] \tag{59}$$

Although only static cases are considered and therefore no time coordinate is specified, in practice each load position $x_{i,2}$ can be associated with a time coordinate t_i. Considering the sensor position as a space coordinate, each point can then be described by a space and time coordinate as shown in Figure 23. The random field is represented as a (not necessarily regular) grid of points in a spatial and a temporal dimension with n denoting the total number of sensors and m denoting the number of observations of each influence line.

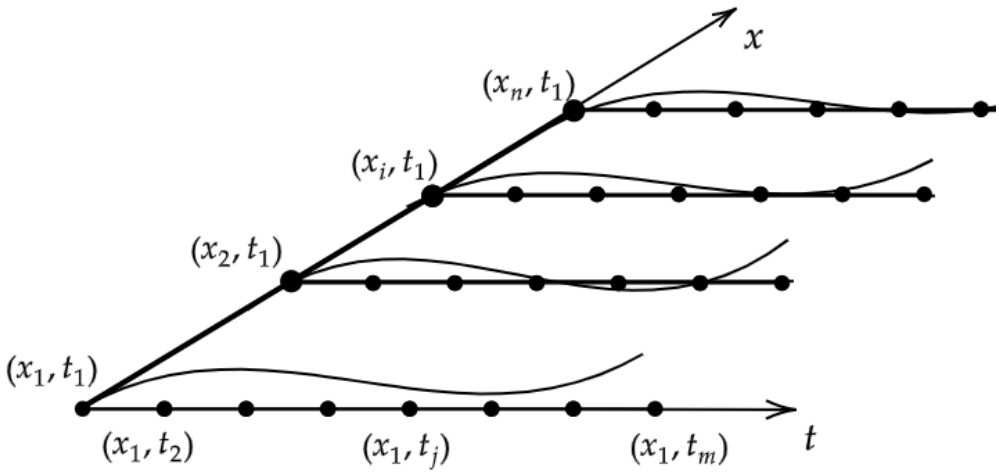


Figure 23: Illustration of space and time coordinate system. Influence lines along the time axis t are obtained for each sensor position x.

This interpretation of the sensor and load position as space and time coordinates is adopted both for convenience in notation and to enable more intuitive explanations of the analyses that follow. The vector $\boldsymbol{x}$ becomes:

$$\boldsymbol{x}_i = [x_i, t_i] \tag{60}$$

The correlation function is multiplied by the standard deviation of the modeling uncertainty to form the covariance function. Calculating the covariance for every pair of points yields the symmetric positive semi-definite covariance matrix shown in Equation 61, that describes the covariance of the model prediction error for every point in the random field.

$$\Sigma_{\text{model}} = \begin{bmatrix} \sigma_1^2 k(\boldsymbol{x_1}, \boldsymbol{x_1}; \boldsymbol{\theta}_\text{c}) & \dots & \sigma_1 \sigma_N k(\boldsymbol{x_1}, \boldsymbol{x_N}; \boldsymbol{\theta}_\text{c}) \\ \vdots & \ddots & \vdots \\ \sigma_N \sigma_1 k(\boldsymbol{x_N}, \boldsymbol{x_1}; \boldsymbol{\theta}_\text{c}) & \dots & \sigma_N^2 k(\boldsymbol{x_N}, \boldsymbol{x_N}; \boldsymbol{\theta}_\text{c}) \end{bmatrix}, \tag{61}$$

where $N = n \cdot m$. It is noted that for the majority of the analyses performed, only a single sensor is taken into account. In that case only the time coordinate is needed to define the location of each measurement and the coordinate vector $\boldsymbol{x}_i$ for point i becomes a scalar t_i, where $i \in [m]$. Each set of measurements for a given sensor can then be represented as a time series $\boldsymbol{t} = [t_1, t_2, ..., t_i, ..., t_m]$. Similar to Equation 61, the covariance matrix is then formulated as:

$$\Sigma_{\text{model}} = \begin{bmatrix} \sigma_1^2 k(t_1, t_1; \boldsymbol{\theta}_\text{c}) & \dots & \sigma_1 \sigma_m k(t_1, t_m; \boldsymbol{\theta}_\text{c}) \\ \vdots & \ddots & \vdots \\ \sigma_m \sigma_1 k(t_m, t_1; \boldsymbol{\theta}_\text{c}) & \dots & \sigma_m^2 k(t_m, t_m; \boldsymbol{\theta}_\text{c}) \end{bmatrix}, \tag{62}$$

In Equation 61 and Equation 62, σ is dependent on the probabilistic model formulation. Under the assumption of additive modeling uncertainty, the marginal variance for each element of the covariance matrix can be calculated as $\sigma_{i,j} = \sigma_i \cdot \sigma_j$. For multiplicative modeling uncertainty the marginal variance is replaced by the product of the coefficient of variation C_v and the model output, calculated as $\sigma_{i,j} = C_{v,i} \cdot y_i \cdot C_{v,j} \cdot y_j$.

5.4.2 Likelihood function

Complete independence is assumed between the right and left lane model prediction uncertainties. For each point sampled from the posterior distribution, the total likelihood is obtained as the product of the likelihoods for each lane:

$$L(\boldsymbol{\theta}) = \prod_{l=1}^{2} L_l(\boldsymbol{\theta}), \tag{63}$$

where the likelihood function is taken as in Equation 11. The subscript $l = 1$ and $l = 2$ denotes the left and right lane influence line respectively, and the terms $\boldsymbol{y}_{\text{meas},l}$ and $\boldsymbol{y}_{\text{model},l}$ denote the vector of observations and model prediction respectively for the l^{th} lane. The vectors $\boldsymbol{y}_{\text{meas},l}$ and $\boldsymbol{y}_{\text{model},l}$ have shape:

$$\boldsymbol{y}_l = [y_{l,1,1}, y_{l,1,2}, ..., y_{l,1,m}, y_{l,2,1}, y_{l,2,2}, ..., y_{l,n,1}, y_{l,n,2}, ..., y_{l,n,m}], \tag{64}$$

and Σ_i is the $N \times N$ covariance matrix.

5.4.3 Probabilistic model parameters to be estimated

Prior distributions of the uncertainty parameters for the considered probabilistic models are provided in Table 8 and apply for all analyses unless otherwise specified. The support (domain) of the priors must be defined such that it is possible to capture the structure of the correlations in the residuals between model and measurements. It should be noted that choosing the priors for the uncertainty parameters is not a simple task as no information on the correlation structure is available. Furthermore, a poor choice of prior can significantly impact the inference and prediction, leading to wide credible intervals (see e.g. Fuglstad et al. (2018)). Uniform priors are chosen with supports that are wide enough to capture a range of correlations that are expected to be present in the measurements.

Table 8: Prior distributions of uncertainty parameters.

Model	C_v	l_corr [m]	w_n	α
Independent	$\mathcal{U}(0.00, 1.00)$	-	-	-
Radial Basis Function	$\mathcal{U}(0.00, 1.00)$	$\mathcal{U}(0, 200)$	-	-
Rational Quadratic	$\mathcal{U}(0.00, 1.00)$	$\mathcal{U}(0, 200)$	-	$\mathcal{U}(0.00, 5.00)$
Damped Cosine	$\mathcal{U}(0.00, 1.00)$	$\mathcal{U}(0, 1000)$	$\mathcal{U}(0.00, 0.50)$	-
Matern $\nu = 1.5$	$\mathcal{U}(0.00, 1.00)$	$\mathcal{U}(0, 200)$	-	-

The coefficient of variation parameter C_v is only applicable for the case of multiplicative modeling uncertainty. In the case of additive modeling uncertainty the standard deviation parameter is used instead, distributed as $\sigma_\mathrm{model} \sim \mathcal{U}(0.00, 20.0)$ MPa for all models. The standard deviation of the measurement uncertainty is taken as $\sigma_\mathrm{meas} = 1.0$ MPa.

5.5 Concluding summary

The IJsselbridge case study is detailed in this section. The approximately 300 m steel bridge is composed of two variable height steel plate girders connected by K-braces, supporting a steel deck with longitudinal bulb stiffeners. Details on two measurement campaigns are provided, describing the controlled loading conditions and sensor layouts used to obtain the data, as well as the post-processing of the data to obtain stress influence lines. Additionally, details on the FE models used to simulate the structural response are provided. The Euler-Bernoulli beam element formulation is used to define a single and a twin girder model, capturing the structural behaviour for a load moving in the longitudinal direction of the bridge and a linear LLF accounts for the transverse position of the load. Depending on the case (see Section 6), this model can be parametrized by the slope coefficient c_1 of the LLF as well as the rotational stiffnesses of the support springs. Uniform prior distributions are considered for these parameters.

Similarly, a description of the probabilistic models is provided. Each influence line is considered as a time series measured at a point along the longitudinal direction of the bridge, and the discrepancies between measurements and model predictions form a Gaussian random field in a spatial and a temporal dimension. The data generating processes described

in Sections 3.2.1 and 3.2.2 are used to formulate the Gaussian likelihood function. Different probabilistic models are defined based on the covariance function used in the formulation and uniform prior distributions are assumed for the parameters of the covariance functions.

6 Analysis of the effect of spatial and temporal dependencies on system identification and sensor placement for the IJsselbridge

6.1 Description of cases

The main research questions posed in Section 1.2 are investigated in this section by application to the IJsselbridge case study. The feasibility of inferring the dependence structure is addressed by formulating a pool of candidate models corresponding to different correlation structures and using them to perform Bayesian inference of both the physical and statistical model parameters. Bayesian model selection is then applied to determine the most suitable statistical model. The analyses presented here are performed using Nested sampling (see Section 3.4) to address the issue of efficiently calculating the evidence. Additionally, a method for performing sensor placement optimization based on the approximation of the information entropy (see Section 3.5.1) is implemented. This method is applied to determine the most informative measurements for the IJsselbridge. Using this measurement selection technique, Bayesian inference is performed for increasing numbers of measurements to determine the effect of the sample size on the posterior distributions of the parameters and the posterior predictive distributions of the stress influence lines. Where necessary, reference will be made to the examples presented in Appendix D to support the analyses and arguments presented here and to determine the applicability of the methods used. In total, three cases are presented:

- **Case 1**: An initial exploratory analysis with a simplified physical model parametrization and data from a single sensor.
- **Case 2**: A more in-depth analysis with additional physical model parameters and data from a single sensor.
- **Case 3**: Extension to multiple sensors and combined spatial and temporal dependencies considering additive model prediction uncertainty.

Each of the cases includes some of the steps described previously. An overview of the cases is provided in Table 9.

Table 9: Overview of sub-cases for the IJsselbridge case study.

	Case 1	Case 2	Case 3
Sensors	PBR-P4	PBR-P4	H1 H4, H5, H9, H10
Modeling uncertainty	Multiplicative	Multiplicative	Additive
Parameters	C_v, l_{corr}, w_n, α, c_1	C_v, l_{corr}, w_n, α, c_1, $K_{r,1}$, $K_{r,2}$, $K_{r,3}$, $K_{r,4}$	σ, l_{corr} c_1,$K_{r,1}$, $K_{r,2}$, $K_{r,3}$, $K_{r,4}$

A shorthand notation will be adopted to refer to the models for the remainder of this thesis. The models are distinguished by the kernel function used in the statistical model and the size of the dataset. The kernels are therefore abbreviated as follows:

- Independence - IID
- Radial basis function - RBF
- Rational quadratic function - RQD
- Damped cosine function - COS
- Matern with $\nu = 1.5$ - MAT
- Exponential - EXP

Models will be denoted as N/IL, Kernel, indicating the number of points per influence line and kernel function. Following this notation, a model with 165 points per influence line using a Matern kernel will be written as: 165/IL,MAT. Alternatively, the index i of the model $\mathcal{M}_i$ when referring to more than two models.

6.2 Case 1: Impact of dependencies in model prediction error

Given that the correlation parameters are not known a-priori and often lack physical meaning, the available data must be used to determine the correlation structure, leading to inefficient use of measurements which would otherwise provide information on the parameters of interest. In this section this issue is addressed by performing system identification for the IJsselbridge case, introduced in Section 5, using a class $\boldsymbol{\mathcal{M}}$ of coupled probabilistic - physical models. Each model $\mathcal{M}_i \in \boldsymbol{\mathcal{M}}$ is constructed with a different kernel function that could potentially describe the dependencies in the model prediction error. The set of uncertain parameters for each model $\mathcal{M}_i$, $\boldsymbol{\theta}_i$ is extended to include the set of uncertainty parameters of the corresponding correlation function $\boldsymbol{\theta}_{\mathrm{c,i}}$, in order to assess the feasibility of simultaneously inferring the structural and correlation parameters in Bayesian system identification. Prior to this analysis, a minimal example is developed to determine the effect of the number of points in a domain on the posterior mean and variance of the correlation length. This can be found in Appendix D.2.

In total, six models are considered. Models $\mathcal{M}_1$ - $\mathcal{M}_5$ are distinguished by the type of kernel used to model the dependencies in the model prediction error: Independence, radial basis, rational quadratic, damped cosine and Matern with $\nu = 1.5$. A reference model using four points per influence line and assuming independence in the model prediction error is also defined. This model represents the typical analysis procedure for Bayesian system identification where sufficiently spaced data points are selected such that the assumption of independence is valid. The dataset used for models $\mathcal{M}_1$ - $\mathcal{M}_5$ contains 165 equally spaced points along the entire influence line. One influence line from sensor PRB-P4 is used per lane, for a total of two. The influence lines for the left and right lane, as well as the selected peaks used in the reference model are shown in Figure 14. The single girder FE physical model is used in

all model formulations, parametrized by the slope coefficient c_1 of the LLF. Pinned supports are considered with zero rotational stiffness. This parametrization is chosen for simplicity since the main objective is the inference of the statistical parameters. A summary of the models for this case is provided in the table below.

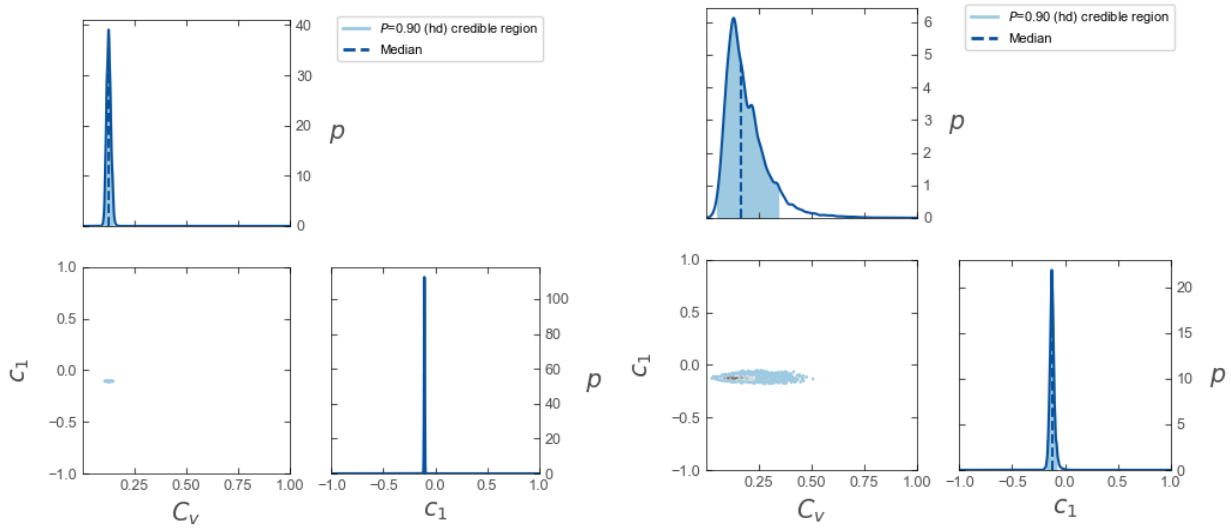
Table 10: Summary of parameters for case 1.

Case #:	1	Sensors:	PRB-P4
Physical model:	Single girder	$\boldsymbol{\theta}_\mathrm{s}$:	c_1
Modeling uncertainty:	Multiplicative	Meas. uncertainty:	Additive, $\sigma_\mathrm{meas} = 1.0$ MPa

Model #	Correlation	Dataset size	$\boldsymbol{\theta}_\mathrm{c}$
1	Independent	165/IL	C_v
2	Radial basis	165/IL	C_v, l_corr
3	Rational quadratic	165/IL	C_v, l_corr, α
4	Damped cosine	165/IL	C_v, l_corr, w_n
5	Matern $\nu = 1.5$	165/IL	C_v, l_corr
6	Independent	4/IL	C_v

6.2.1 Inferring the dependence structure

Inference is performed using the nested sampling technique described previously for the models presented in Table 10. A range of approximately $2.1 \cdot 10^4 - 3.3 \cdot 10^4$ likelihood evaluations are needed to achieve convergence depending on the model. The marginal and joint posterior distributions of the parameters for the 165/IL, IID and 4/IL, IID are shown in Figure 24. It is noted that in the 165/IL,IID model, complete independence is assumed despite the dense spacing of the measurements and it is expected that the uncertainty in the posterior distributions will be significantly underestimated if dependencies are present in the model prediction error. The posterior distributions obtained by this model are therefore included for the purpose of comparing the inferred means of the parameters with other models, as well as illustrating the effect of increasing the number of points under the independence assumption on the posterior distributions.



(a) 165/IL, IID (b) 4/IL, IID

Figure 24: Comparison of marginal and joint posterior distributions of parameters for the 165/IL,IID and 4/IL,IID models.

Comparing the posteriors of the 4/IL,IID and 165/IL,IID models, it is evident that the posterior parameter uncertainty is reduced for larger N. Both the median point estimate and the standard deviation of the modeling uncertainty coefficient of variation C_v are lower for this model. The reduction in uncertainty for the 165/IL, IID model can also be observed in the posterior c_1, although to a lesser extend. Summary statistics of the posterior distributions for the two models are provided in Table 11.

Table 11: Mean, median, std. dev and percentiles of posterior distributions for the 165/IL,IID and 4/IL,IID model.

Parameter	Model	Mean	Median	Std. dev.	p25	p75	p95
C_v	1	0.12	0.12	0.01	0.12	0.13	0.14
	6	0.19	0.17	0.10	0.12	0.24	0.38
c_1	1	-0.10	-0.10	0.00	-0.11	-0.10	-0.10
	6	-0.12	-0.12	0.02	-0.14	-0.11	-0.08

Posterior distributions of the parameters for models 2, 3, 4, and 5 are shown in Figure 25. Although the correlation length l_corr is a common parameter for these models and is inferred for the radial basis, rational quadratic and Matern models, it can be seen that different point estimates and uncertainties are obtained. Notably, the damped cosine model exhibits a wide posterior for the correlation length with little reduction of the uncertainty compared to the specified $\mathcal{U}(0.0, 1000.0)$ prior distribution. Additionally, the joint distributions of C_v and l_corr indicates that the two parameters are correlated. This issue of unidentifiability has been previously mentioned in literature, e.g. in Fuglstad et al. (2015), and no single

solution can be found in the 1D case. This results in wider marginal posteriors for the two parameters and potentially higher point estimates for the modeling uncertainty. Finally, the wide tails observed in the 165/IL,MAT and 165/IL,RQD posteriors for C_v and l_{corr} highlight the difficulty of inferring those parameters.

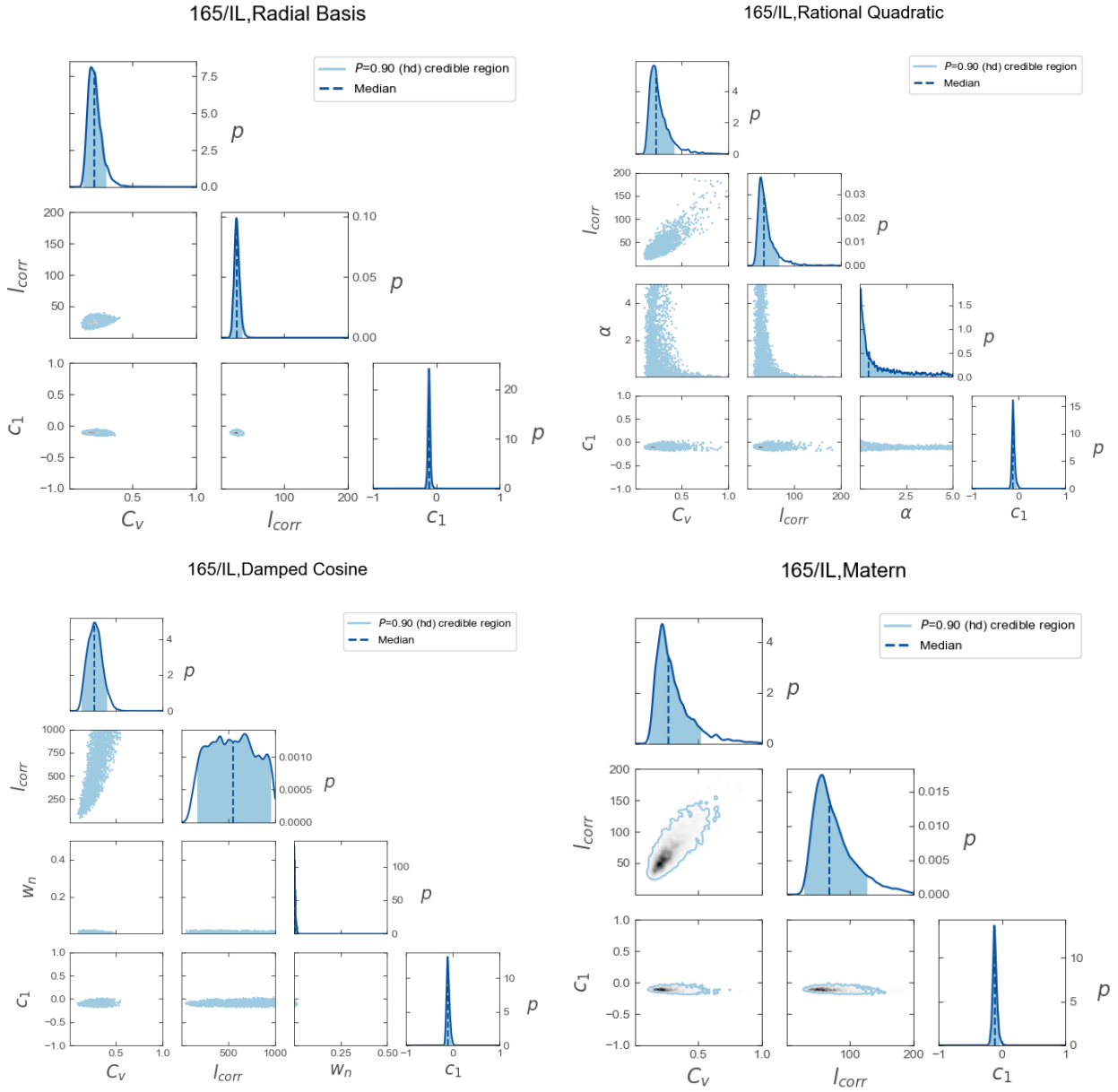


Figure 25: Comparison of posterior distributions for models $\mathcal{M}_2$-$\mathcal{M}_5$.

A summary of the posterior distributions for the full class of models is provided in the table below. A comparison of the models that include correlation indicates that the 165/IL,RBF model achieves the lowest uncertainty in the posterior distribution of c_1. Similarly, the lowest posterior mean for C_{v} is obtained for the RBF model, with the remaining models resulting in higher point estimates for both C_{v} and l_{corr}. This variability between estimates is attributed to the different formulation of the correlation for each model, as well as the different number

of parameters. The latter is discussed in detail in the following sections, however, it is generally observed that increasing the number of inferred parameters for constant dataset size N results in higher posterior uncertainty. The means and coefficients of variation (COV[7]) of the uncertainty parameters are shown in Table 12.

Table 12: Means of posterior distributions for case 1 and COV in parenthesis.

Model	C_v [-]	l_corr [m]	α [-]	w_n [-]	c_1 [-]
165/IL,IID	0.125 (0.083)	-	-	-	-0.105 (0.034)
165/IL,RBF	0.203 (0.276)	25.100 (0.175)	-	-	-0.112 (0.172)
165/IL,RQD	0.261 (0.444)	40.914 (0.496)	1.109 (1.177)	-	-0.107 (0.309)
165/IL,COS	0.274 (0.281)	547.542 (0.443)	-	0.005 (0.821)	-0.098 (0.342)
165/IL,MAT	0.305 (0.481)	75.214 (0.432)	-	-	-0.112 (0.309)
4/IL,IID	0.193 (0.542)	-	-	-	-0.121 (0.196)

In order to accurately quantify the uncertainty in the predicted stress influence lines, all of the components contributing to the uncertainty must be taken into account. The posterior predictive distributions of the stress influence lines are calculated, yielding a stress distribution $p(\bar{\boldsymbol{y}}_i|\boldsymbol{y}_\mathrm{meas})$ at each load position i which contains the combined measurement, modeling and parameter uncertainty. The obtained credible intervals (CIs) for the measurement and combined uncertainty are also indicated in the plots. In the following figure, the posterior predictive distribution for the 4/IL,IID reference model is compared with the 165/IL,IID model.

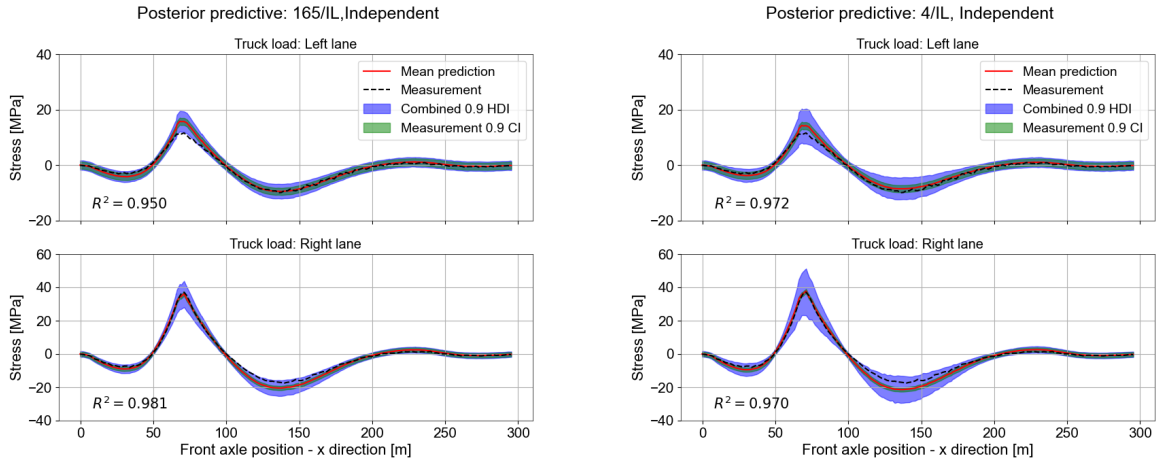


Figure 26: Comparison of posterior predictive distributions of the stress influence lines for models 1 and 6.

For the 165/IL,IID model is is evident that the width of the combined uncertainty CI corresponds to the maximum discrepancy between model and measurement. The large N and

[7]The coefficient of variation of the inferred parameters will be denoted as COV to distinguish it from the modeling uncertainty coefficient of variation (C_v) parameter.

assumption of independence in the residuals results in overconfidence in the posterior parameter distributions and underestimation of the uncertainty. In contrast, for the 4/IL,IID model there is significant uncertainty in the estimation of C_v, which is reflected in the wide CIs. The degree of model fit to the data, quantified by the R^2 score is similar for both models. Given the difference in dataset size, this indicates that the physical model fit to the data can not be further improved under the current parametrization. A comparison of the models with correlation is shown below.

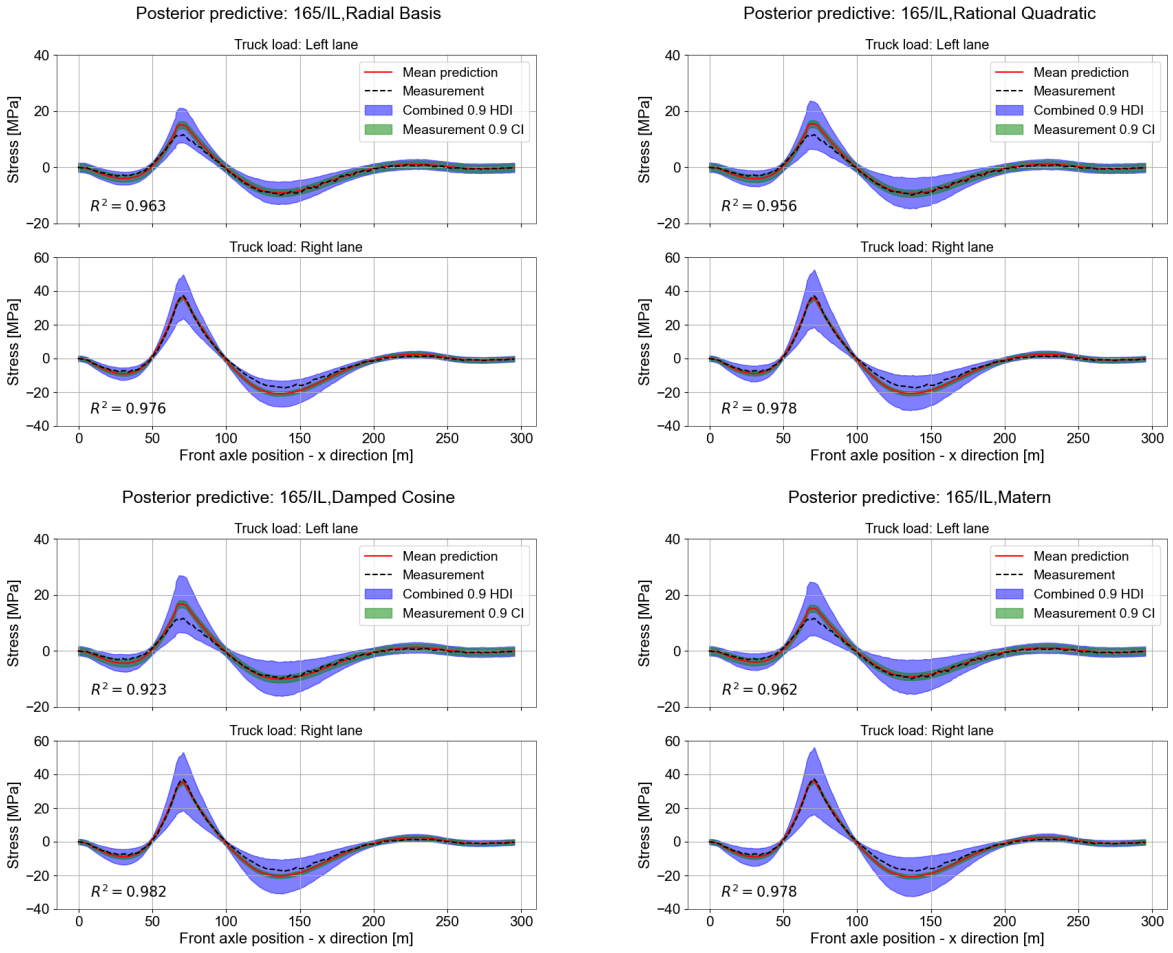


Figure 27: Comparison of posterior predictive distributions of the stress influence lines for models 2, 3, 4 and 5.

It is evident that the unidentifiability of the marginal variance and correlation length results in wide CIs for models $\mathcal{M}_2$-$\mathcal{M}_5$. To quantify the model fit to the measurements, the R^2 score, mean squared error (MSE), maximum error (ME) and error at the peak (PE) are calculated. For this case, the metrics are largely inconclusive as the models yield similar measures of fit and additionally a trade off between the left and right lane fit is generally observed. The metrics of fit for each model are summarized in Table 13.

Table 13: R2 score, mean squared error, max. error and error at peak for case 1.

Case 1	R^2 [-]		MSE [MPa^2]		ME [MPa]		PE [MPa]	
Model	Left	Right	Left	Right	Left	Right	Left	Right
1	0.95	0.98	1.22	2.51	4.62	3.90	4.15	0.81
2	0.96	0.98	0.83	3.17	3.98	4.26	3.98	0.08
3	0.95	0.98	0.99	2.44	4.84	4.02	4.84	0.38
4	0.93	0.98	1.70	1.99	5.49	3.52	5.49	1.84
5	0.97	0.98	0.81	3.49	3.63	4.43	3.20	0.43
6	0.97	0.97	0.65	4.02	3.21	4.81	3.00	0.73

Taking into account the correlations in the modeling uncertainty reduces the effective sample size, resulting in wider CIs. For this case the correlated models result in similar or higher prediction uncertainty compared to the reference model. Intermediate conclusions from this case are summarized below.

- The uncertainty parameters describing the correlation structure can be inferred from the data, however, the unidentifiability of C_v and l_{corr} results in high prediction uncertainty.
- Considering correlations reduces the effective sample size.
- Larger N generally results in more accurate inference and prediction.
- Simple parametrization limits the conclusions that can be drawn from this case. This analysis will therefore be repeated for extended physical model parametrization.
- No significant improvement is observed in terms of model fit or prediction uncertainty for the models with dependencies compared to the reference model.

6.2.2 Calculation of the evidence and model selection

The structure of the correlation, and thus the probabilistic model that can best describe it, are not known a priori and are inferred from the available observations by application of Bayesian model selection using the calculated evidence. As discussed previously in Section 3.1.2, calculating the evidence by numerical integration of the denominator in Equation 1 is not possible in case of a high dimensional parameter space (e.g. for $N_\theta > 3$) where N_θ is the number of parameters. In this section the feasibility of the following is investigated:

- Utilizing the nested sampling technique to obtain an estimate of the evidence.
- Applying Bayesian model selection to determine the model that best describes the dependencies in the model prediction error.

Nested sampling is used to obtain estimates of the evidence for the coupled probabilistic-physical models defined in Section 6.2. It is emphasized that model selection is dependent

on the set of observations $\boldsymbol{y}_{\text{meas}}$ as shown in Equation 2. It follows that only models with the same set of observations can be properly compared using Bayesian model selection. Since the reference model $\mathcal{M}_6$: 4/IL,IID only uses a subset of the observations $\boldsymbol{y}_{\text{meas}}$ used in $\mathcal{M}_1$ through $\mathcal{M}_5$, it can not be included in the selection.

For each model the evidence $\mathcal{Z}$ is calculated. Assuming equal prior probabilities $p_0(\mathcal{M}_i) = 0.2$ for all models and substituting into Equation 2 yields the posterior probabilities $p(\mathcal{M}_i)$. The Bayes factor K_i is calculated by applying Equation 3 with $\mathcal{M}_1$ being the model with the highest probability and $\mathcal{M}_2$ the model i. The calculated log-evidence, posterior probability and Bayes factor[8] for each model, as well as the the interpretation of the Bayes factor are summarized in Table 14.

Table 14: Log-evidence, posterior probability and Bayes factors per model.

Model	$\log(\mathcal{Z})$	$\text{p}(\mathcal{M})$	K	Interpretation
165/IL,IID	-497.94	0.00	1.13E+56	Decisive
165/IL,RBF	-371.07	0.06	8.98E+00	Substantial
165/IL,RQD	-369.10	0.42	1.26E+00	Barely worth mentioning
165/IL,COS	-375.03	0.00	4.74E+02	Decisive
165/IL,MAT	-368.87	0.52	1.00E+00	Barely worth mentioning

It can be readily observed from the previous table that despite having fewer parameters, the 165/IL,IID model has significantly lower posterior probability and is the least likely among the candidate models. This indicates that the assumption of independence does not provide a suitable description of the data. The low posterior probability and high Bayes factor of the 165/IL,COS model can be attributed to the poor model fit that was observed, as well as the additional parameter compared to the RBF and MAT models resulting in additional complexity which is penalized in Bayesian model selection as discussed in Section 3.1.2.

Regarding the model selection results among the correlated models, additional insight is offered in Simoen et al. (2013) where it is illustrated that for a small number of samples (i.e. influence lines) additional investigation is required to determine the most suitable model. This is due to the model selection favouring models with wider posteriors and higher uncertainty among correlation models that can provide a suitable description of the data. It is also shown that for larger numbers of samples, the correct correlation model will be selected. This is not investigated further in this section as the aim is to determine if the correlation structure must be taken into account in the probabilistic model formulation.

6.3 Case 2: Impact of nuber of parameters and sample size

In the analysis of case 1, a simplified parametrization of the physical model was used considering only the LLF coefficient c_1 as the inferred parameter. The calculated posterior

[8]When calculated for multiple models the Bayes factor K_i quantifies the strength of the evidence in favour of the most likely model compared to model i.

distributions and posterior predictive distributions indicate that the structure of the correlation and the uncertainty parameters for different correlation models can be inferred using Bayesian system identification. However, a simple parametrization means that accurate point estimates and reduced uncertainty can be achieved with relatively few measurements, e.g. using four peaks per influence line. With the exception of reduced uncertainty in the estimate of C_v, no improvement in the model fit or posterior parameter uncertainty was observed for the models that consider dependencies compared to the reference case. It is expected that additional physical model parameters will be required to more accurately capture the structural behaviour. Therefore, to improve model fit and evaluate the benefit of considering large N in higher dimensional (in terms of the number of parameters) cases, the rotational spring stiffnesses of supports at the four leftmost pillars F, G, H and J (see Figure 14) are treated as uncertain parameters. A summary of the parameters and models considered in this case is provided in Table 15.

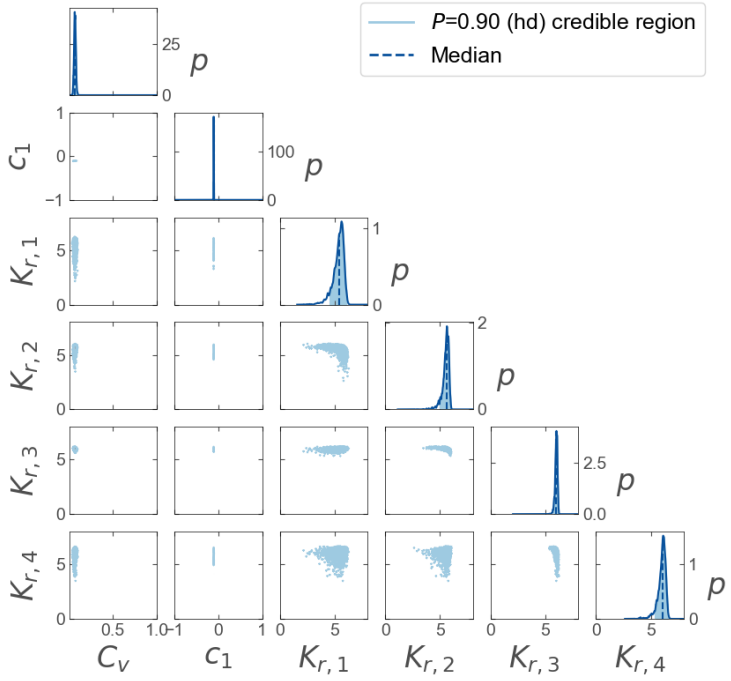
Table 15: Summary of parameters for case 2.

Case #:	2	Sensors:	PRB-P4
Physical model:	Single girder	$\boldsymbol{\theta}_\mathrm{s}$:	$c_1, K_{\mathrm{r}1}, K_{\mathrm{r}2}, K_{\mathrm{r}3}, K_{\mathrm{r}4}$
Modeling uncertainty:	Multiplicative	Meas. uncertainty:	Additive, $\sigma_\mathrm{meas} = 1.0$ MPa

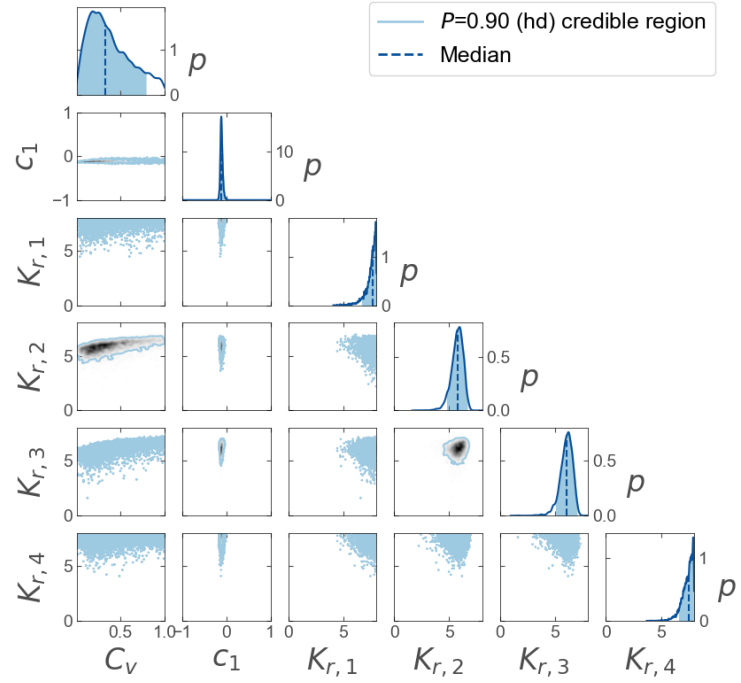
Model #	Correlation	Dataset size	$\boldsymbol{\theta}_\mathrm{c}$
1	Independent	165/IL	C_v
2	Radial basis	165/IL	C_v, l_corr
3	Rational quadratic	165/IL	C_v, l_corr, α
4	Damped cosine	165/IL	C_v, l_corr, w_n
5	Matern $\nu = 1.5$	165/IL	C_v, l_corr
6	Independent	4/IL	C_v

6.3.1 Impact of considering additional parameters

Similarly to case 1, inference is performed using the nested sampling technique for the coupled probabilistic-physical models $\mathcal{M}_1 - \mathcal{M}_6$ with $\mathcal{M}_1$ considering complete independence in the modeling uncertainties, $\mathcal{M}_2$ - $\mathcal{M}_5$ assuming dependencies modeled by different kernel functions and $\mathcal{M}_6$ being the reference model. The uniform prior distributions assumed for the physical model parameters and uncertainty parameters are listed in Table 7 and Table 8 respectively. A comparison of the posterior distributions for models $\mathcal{M}_1$ and $\mathcal{M}_6$ is shown in Figure 28. Note that in the following, the posterior distributions, mean values and standard deviations for the rotational stiffness parameters K_ri are given as $\log_{10}(K_\mathrm{ri})$.



(a) 165/IL, IID



(b) 4/IL, IID

Figure 28: Comparison of marginal and joint posterior distributions of parameters for models 1 and 6. Rotational stiffness plotted in log_{10} scale.

It should be noted that for the 4/IL,IID reference case the posteriors of $K_{\mathrm{r},1}$ and $K_{\mathrm{r},4}$ have wide credible regions and no reduction of uncertainty compared to the uniform prior. This can be seen by plotting the posterior in linear scale:

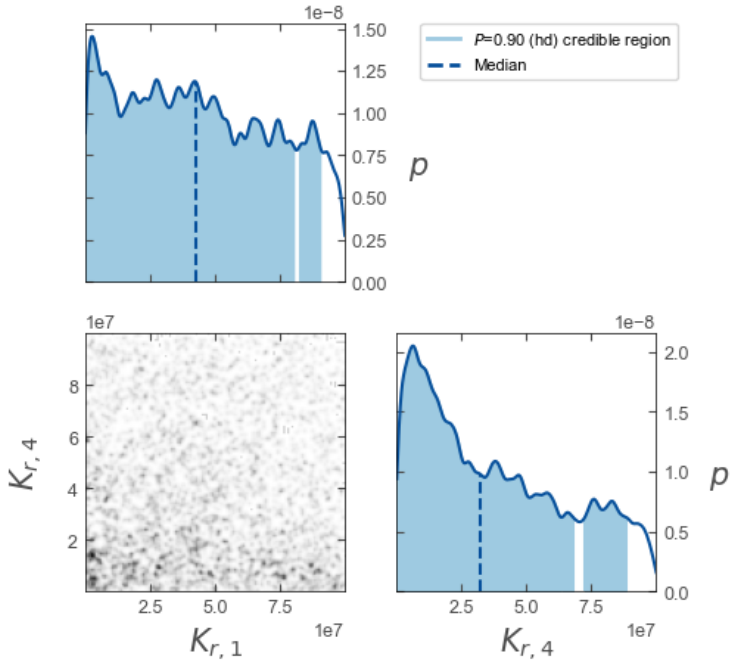
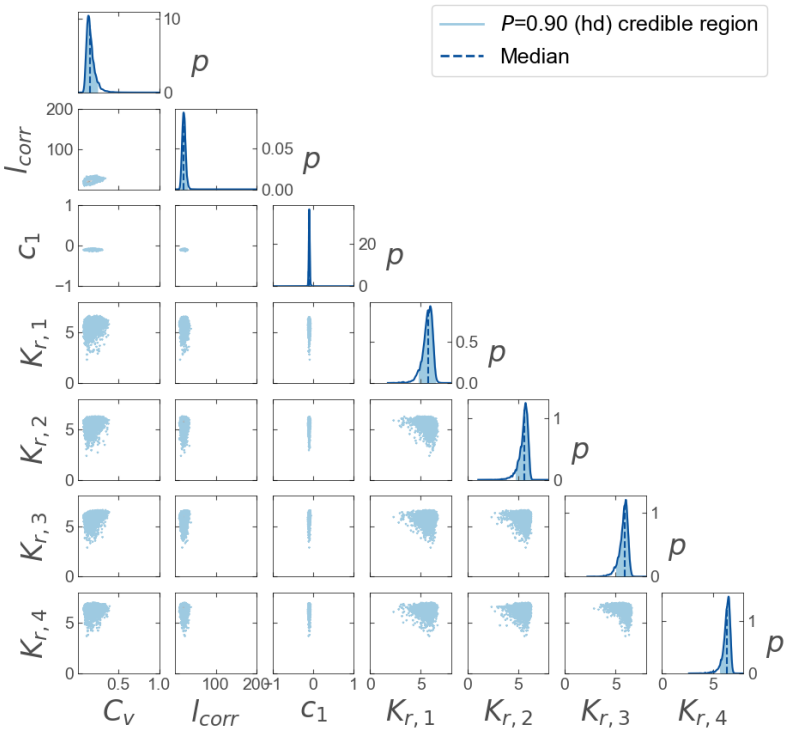
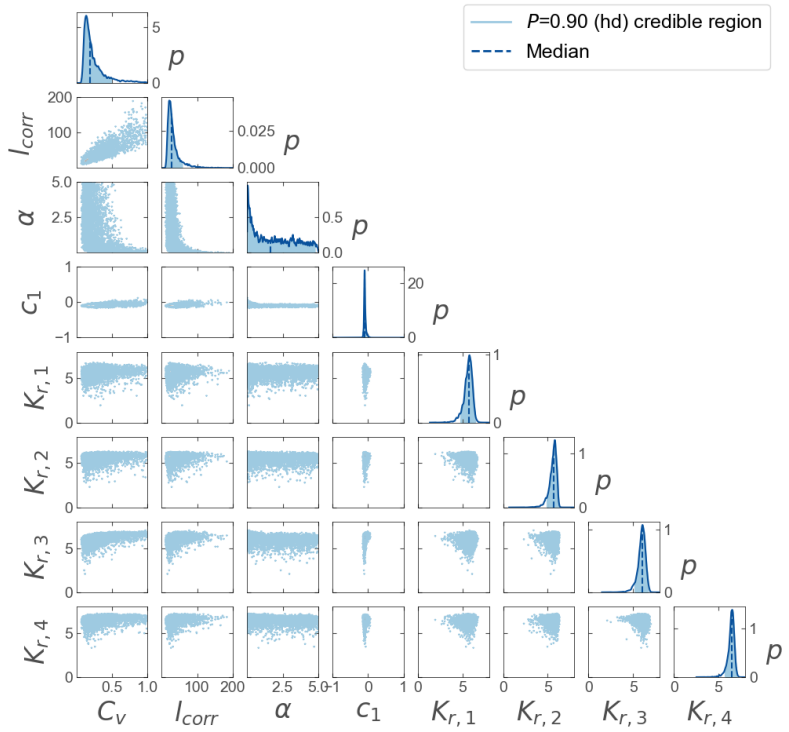


Figure 29: Posterior distribution of $K_{\mathrm{r},1}$ and $K_{\mathrm{r},4}$ in linear scale for the reference model.

It is evident that the Bayesian inference for the 4/IL, IID model results in no reduction of the uncertainty for C_{v}, $K_{\mathrm{r}1}$ and $K_{\mathrm{r}4}$. In contrast these parameters are inferred for the 165/IL,IID model, with the additional data resulting in narrow CR in the posteriors. Although this is expected due to the large number of measurements used and the assumption of independence, it indicates that it is the lack of available data, combined with the wide uniform priors assumed for all parameters that results in the wide CR for the 4/IL,IID model. The posterior distributions for models $\mathcal{M}_2$ through $\mathcal{M}_5$ are shown in the figures below.

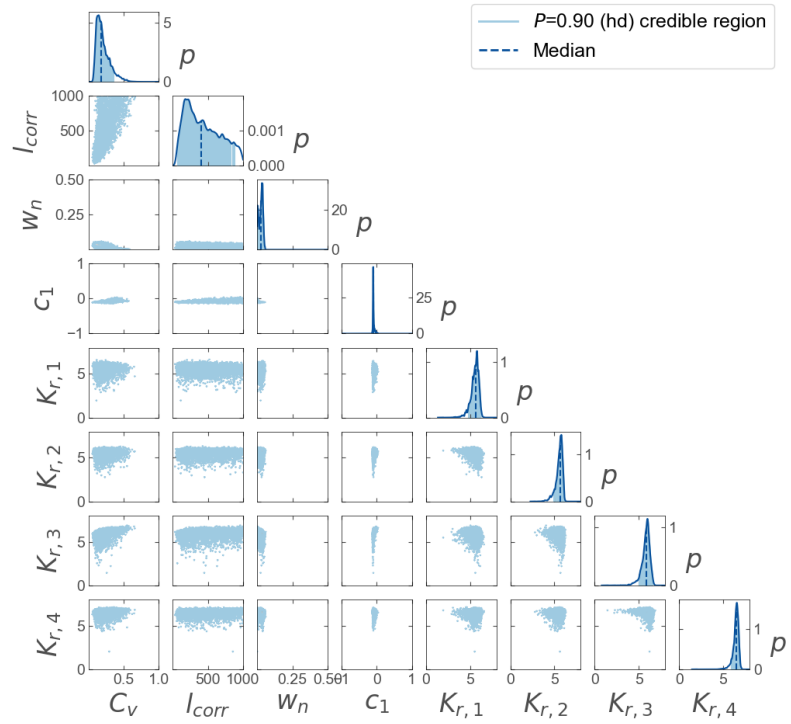


(a) 165/IL, RBF

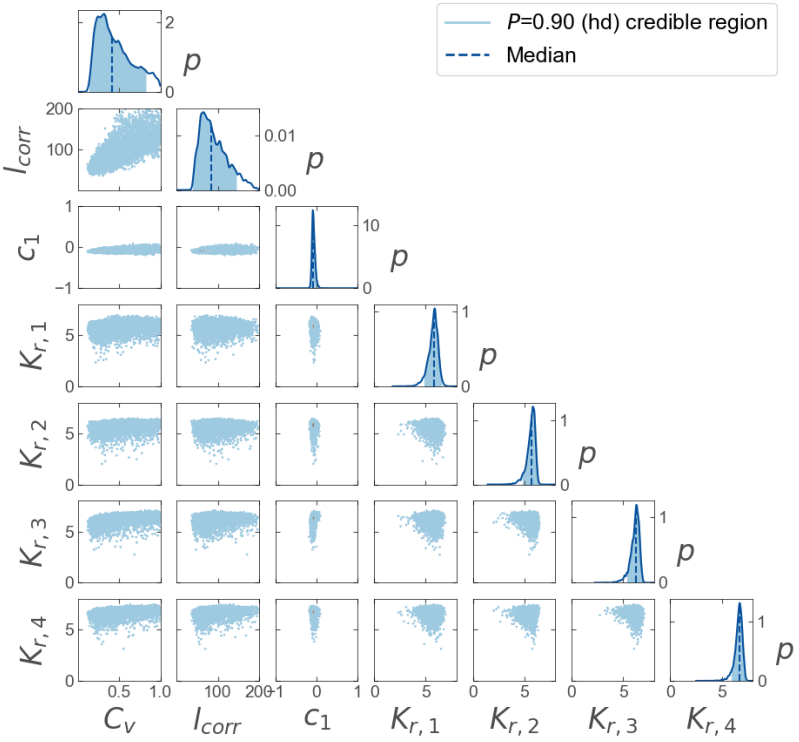


(b) 4/IL, RQD

Figure 30: Comparison of marginal and joint posterior distributions of parameters for models 2 and 3. Rotational stiffness in log scale.



(a) 165/IL, COS



(b) 4/IL, MAT

Figure 31: Comparison of marginal and joint posterior distributions of parameters for models 4 and 5. Rotational stiffness in log scale.

The marginal distributions of the structural parameters c_1, $K_{\mathrm{r}1}$, $K_{\mathrm{r}2}$, $K_{\mathrm{r}3}$ and $K_{\mathrm{r}4}$ show a significant reduction in parameter uncertainty. Furthermore for the 165/IL,RBF a significant reduction of the posterior variance is achieved for the uncertainty parameters C_{v} and l_{corr}. As noted previously, the unidentifiability of C_{v} and l_{corr} is evident from the plots of their joint distributions for all correlated models. The means and COV of the uncertainty parameters are provided in Table 16.

Table 16: Inferred means of uncertainty parameters for case 2 and COV in parenthesis.

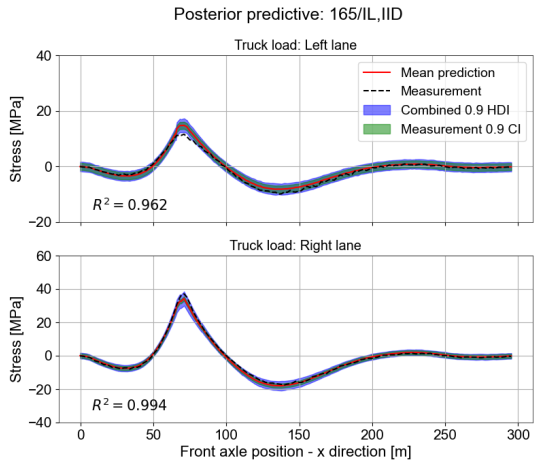
Model	C_{v} [-]	l_{corr} [m]	α [-]	w_{n} [-]
165/IL,IID	0.07 (0.15)	-	-	-
165/IL,RBF	0.16 (0.37)	21.16 (0.20)	-	-
165/IL,RQD	0.25 (0.66)	33.16 (0.58)	1.90 (0.80)	-
165/IL,COS	0.21 (0.47)	439.57 (0.57)	-	0.03 (0.54)
165/IL,MAT	0.46 (0.47)	91.45 (0.36)	-	-
4/IL,IID	0.38 (0.63)	-	-	-

The values of C_{v} indicate that the inferred modeling uncertainty for the 4/IL,IID case is significantly higher. Additionally, the means and COV for the physical model parameters are given in Table 17:

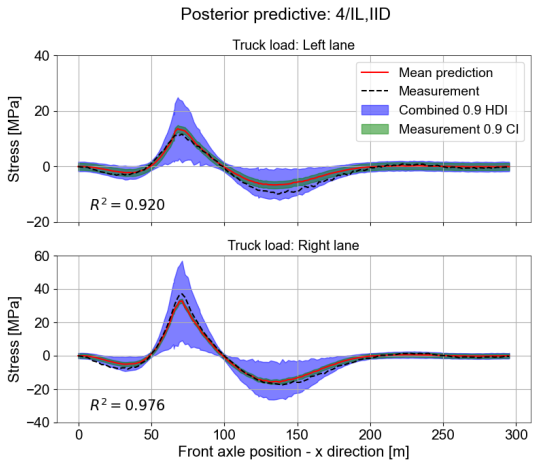
Table 17: Inferred means of physical model parameters for case 2 and COV in parenthesis. Stiffness in kNm/rad.

Model	c_1 [-]	$\log_{10}(K_{r1})$	$\log_{10}(K_{r2})$	$\log_{10}(K_{r3})$	$\log_{10}(K_{r4})$
165/IL,IID	-0.11 (0.02)	5.33 (0.10)	5.50 (0.06)	5.98 (0.02)	5.99 (0.06)
165/IL,RBF	-0.10 (0.13)	5.65 (0.09)	5.53 (0.08)	5.79 (0.08)	6.35 (0.06)
165/IL,RQD	-0.09 (0.33)	5.64 (0.09)	5.57 (0.08)	5.95 (0.08)	6.41 (0.06)
165/IL,COS	-0.09 (0.25)	5.55 (0.08)	5.54 (0.07)	5.80 (0.08)	6.44 (0.05)
165/IL,MAT	-0.08 (-0.54)	5.72 (0.09)	5.63 (0.08)	6.14 (0.07)	6.62 (0.06)
165/IL,IID	-0.11 (-0.29)	7.49 (0.07)	5.72 (0.10)	6.01 (0.10)	7.39 (0.07)

It can be seen that the 165/IL,RBF and 165/IL,RQD models result in lower posterior uncertainty for all parameters compared to the reference model 4/IL,IID. This is also reflected in the posterior predictive distributions shown below. In Figure 32 the 165/IL,IID and 4/IL,IID models are shown. It can be seen that the lack of data in the latter case results in wide CIs. This is the result of high uncertainty in the posterior distribution of C_v (Figure 28a). As previously, the 165/IL,IID model results in narrow CRs for both the posterior and posterior predictive. For this model the CRs generally correspond to the discrepancy between the mean model prediction and measurements. The 165/IL,IID model results also illustrate that increasing the number of available data points results in reduced prediction uncertainty. The effects of increasing the number of measurements used in the inference is further investigated in Section 6.3.3.



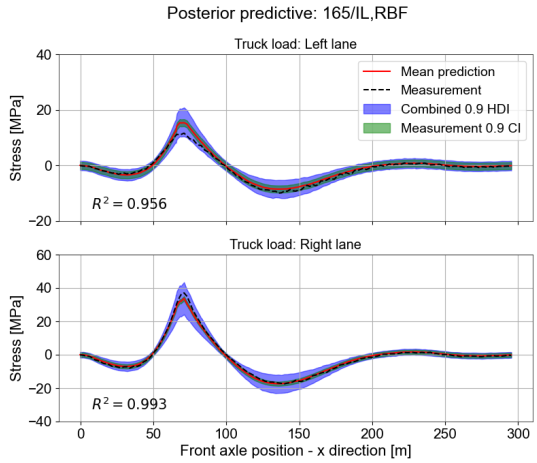
(a) Posterior predictive for $\mathcal{M}_1$.



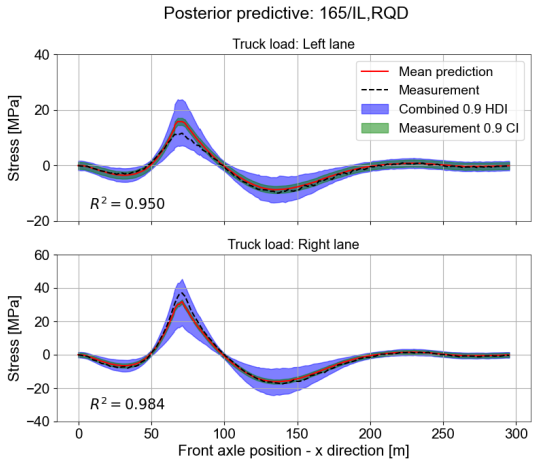
(b) Posterior predictive for $\mathcal{M}_6$.

Figure 32: Posterior predictive distributions of stress influence lines for models $\mathcal{M}_1$ and $\mathcal{M}_6$.

It is expected that taking correlations into account will results in wider CIs compared to the 165/IL,IID model as the effective sample size is reduced. This is confirmed by the posterior predictive distributions for the 165/IL,RBF and 165/IL,RQD models plotted in Figure 33. The lower uncertainty and narrower CIs observed previously for $\mathcal{M}_2$ compared to the other correlated models is also evident in this case.



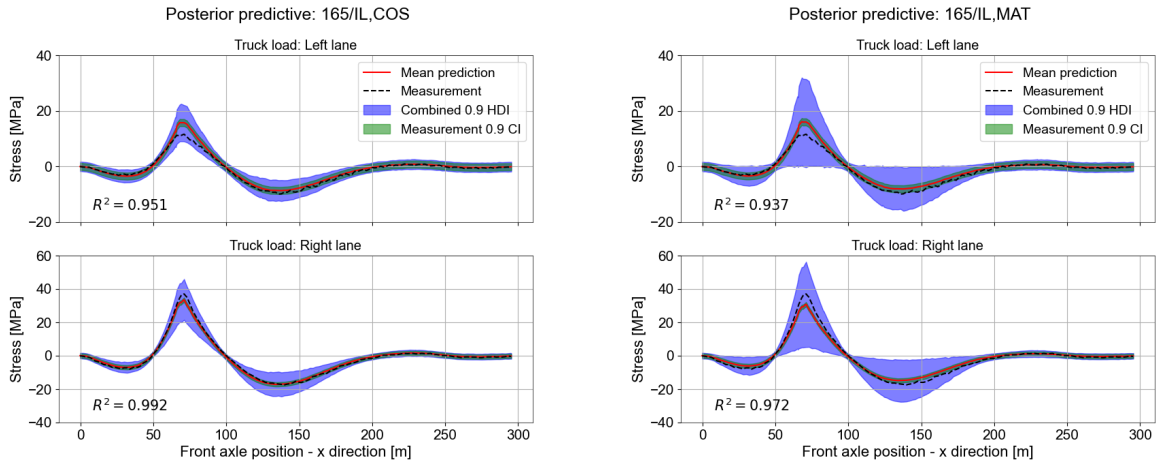
(a) Posterior predictive for $\mathcal{M}_2$.



(b) Posterior predictive for $\mathcal{M}_3$.

Figure 33: Posterior predictive distributions of stress influence lines for models $\mathcal{M}_2$ and $\mathcal{M}_3$.

For models $\mathcal{M}_4$ and $\mathcal{M}_5$, included here for completeness, the high uncertainty in the posterior of the correlation parameters results in wide CIs as shown in the figures below.



(a) Posterior predictive for $\mathcal{M}_4$.

(b) Posterior predictive for $\mathcal{M}_5$.

Figure 34: Posterior predictive distributions of stress influence lines for models $\mathcal{M}_4$ and $\mathcal{M}_5$.

The obtained metrics for the set of models are summarized in the table below:

Table 18: R2 score, Max. error and error at peak for case 2.

Case 1	R^2 [-]		MSE [MPa^2]		ME [MPa]		PE [MPa]	
Model	Left	Right	Left	Right	Left	Right	Left	Right
165/IL,IID	0.96	0.99	0.91	0.78	3.82	3.87	3.82	2.73
165/IL,RBF	0.95	0.99	1.11	1.04	4.41	4.72	3.91	3.61
165/IL,RQD	0.94	0.99	1.43	1.77	5.25	5.70	5.25	4.65
165/IL,COS	0.95	0.99	1.27	1.29	4.94	5.16	4.94	4.14
165/IL,MAT	0.93	0.96	1.66	4.78	5.64	8.15	5.64	7.33
4/IL,IID	0.92	0.98	1.82	3.01	3.32	5.41	2.15	3.84

The advantage in terms of the overall model fit is highlighted by the high MSE values obtained for the 165/IL,RBF model over the reference model. This is not the case for the maximum and peak errors where, as expected, there is no indication of better fit for the RBF model despite the significantly lower modeling uncertainty. These results correspond to expectations: Using only a few peaks in the inference will results in improved fit at these locations at the cost of the overall model fit, whereas using the entire influence line will likely result in a better model fit on average.

The posterior probabilities for each model obtained by performing Bayesian model selection are provided in the table below. As for the previous case, the posterior probability of the 165/IL,IID model is significantly lower compared to the correlated models.

Table 19: Log-evidence, posterior probability and Bayes factors per model for case 2.

Model	$\log(\mathcal{Z})$	$\mathrm{p}(\mathcal{M})$	K	Interpretation
165/IL,IID	-450.74	0.00	3.55E+30	Decisive
165/IL,RBF	-381.18	0.24	2.19E+00	Barely worth mentioning
165/IL,RQD	-380.40	0.53	1.00E+00	Barely worth mentioning
165/IL,COS	-385.01	0.01	1.00E+02	Decisive
165/IL,MAT	-381.30	0.22	2.45E+00	Barely worth mentioning

Based on the results presented throughout this section we can conclude that one of the benefits of considering large N with correlations is the possibility of increasing the number of physical model parameters, resulting in better fit and smaller modeling uncertainty. This could also imply that there is a limit to the number of parameters that can be inferred under the independence assumption, due to the limited number of measurements that can be selected from the influence line while neglecting the correlations between them. Furthermore, it is clear that considering a small number measurements at the peaks will generally result in a better model fit to those peaks, while a large number of measurements distributed along the length of the bridge will provide a better overall fit.

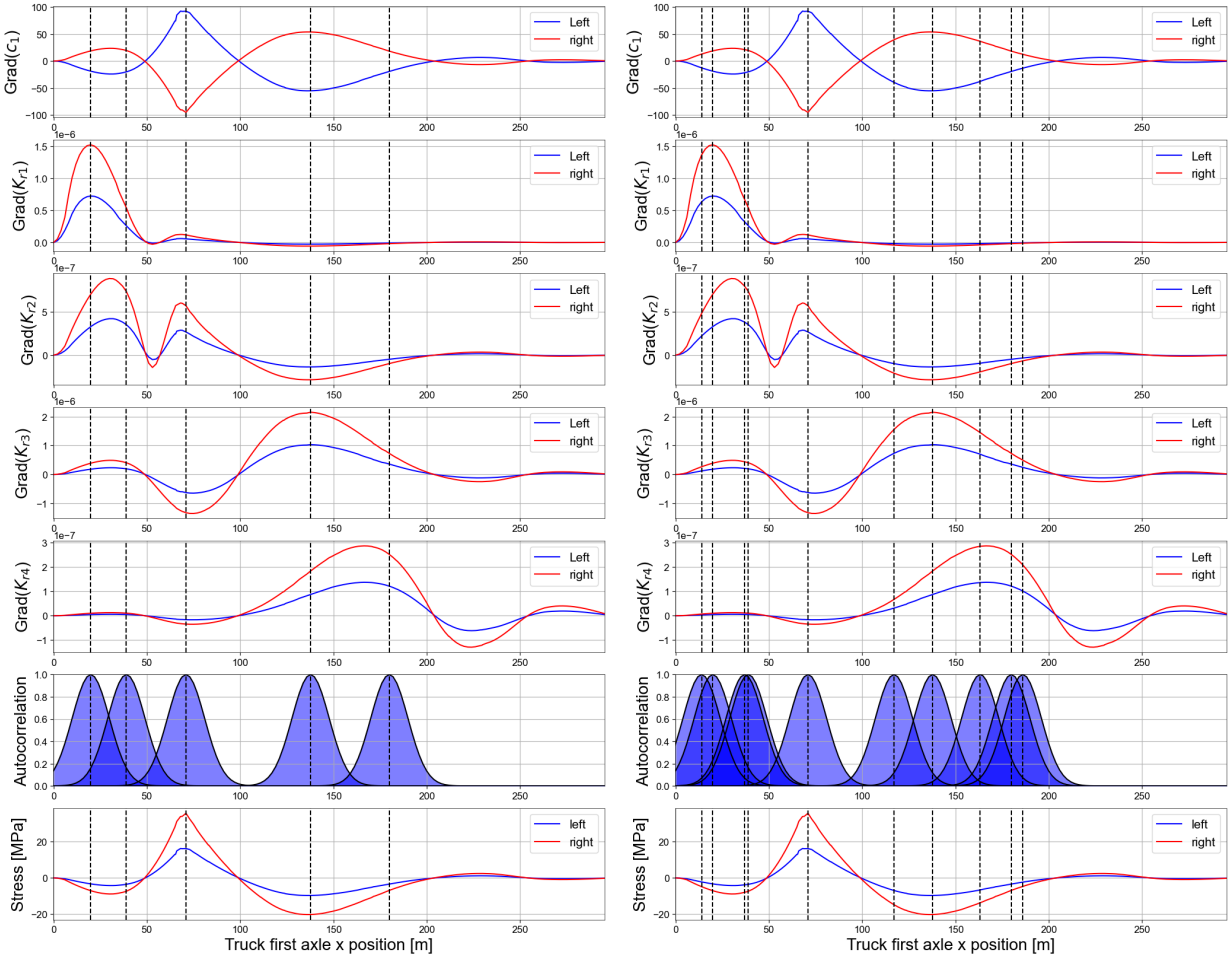
6.3.2 Selection of informative measurements

The sensor placement optimization method developed by Papadimitriou and Lombaert (2012), described in Section 3.5.1 is implemented in order to evaluate the applicability for the case of the IJsselbridge. A thorough comparison of the approximate method with an exact approach using the KL-divergence is presented in Appendix D.5. Given that the computational cost of the likelihood evaluation typically scales with the cube of the number of points, it becomes important when a large number of measurements is available to select a subset to use in inference. Mathematically, this problem is similar to determining the set of sensors that provide the most informative measurements. This method enables the selection of the measurements with the highest information content and would result in the largest reduction in posterior uncertainty for the parameters of interest when used in system identification.

Measurement selection is applied for the single girder IJsselbridge model considering all of the structural parameters, i.e. $\boldsymbol{\theta}_s = \{c_1, K_{\mathrm{r}1}, K_{\mathrm{r}2}, K_{\mathrm{r}3}, K_{\mathrm{r}4}\}$. The nominal values of the physical and uncertainty parameters are taken as follows:

- $c_1 = -0.1$
- $K_{\mathrm{r}1-4} = 10^4$ kNm/rad
- $C_{\mathrm{v}} = 0.1$
- $\sigma_{\mathrm{meas}} = 1.0$ MPa
- $l_{\mathrm{corr}} = 10.0$ m

The figures below depict the gradient of the left and right truckload influence lines w.r.t. each of the structural parameters, as well as the normalized autocorrelation at the location of each measurement and the stress influence lines. Only results for $N = 5$ and $N = 10$ are shown for clarity.



(a) Measurement selection for $N = 5$.

(b) Measurement selection for $N = 10$.

Figure 35: Measurement selection for the IJsselbridge case for $N = 5$ and $N = 10$ measurements.

It can be seen that the selected measurements generally correspond to the peak locations of the gradients. Some clustering of the sensors is observed at the first span located in the region $x = [0.0, 50.0]$ m. This is attributed to the high information content in that region that is evident by the high values of the gradients for c_1, K_{r1} and K_{r2} and the spacing of the measurements imposed by the assumption of correlation. Generally, the resulting selection is compatible with expectations and is applied when studying the effect on the inference of increasing the number of measurements.

6.3.3 Effect of increasing number of measurements

In order to determine if accurate inference for the IJssel bridge can be performed with a few manually selected measurements, and to determine the benefit of considering large N, the posterior mean and standard deviation of the unknown parameters as a function of N are calculated. This is done by performing Bayesian inference for different values of N in the range $[4, 165]$. The most informative measurements for each N are then selected using the procedure detailed in 6.3.2. The selected measurement locations are plotted as a function of N in the figure below, with the red dashed lines denoting the support locations.

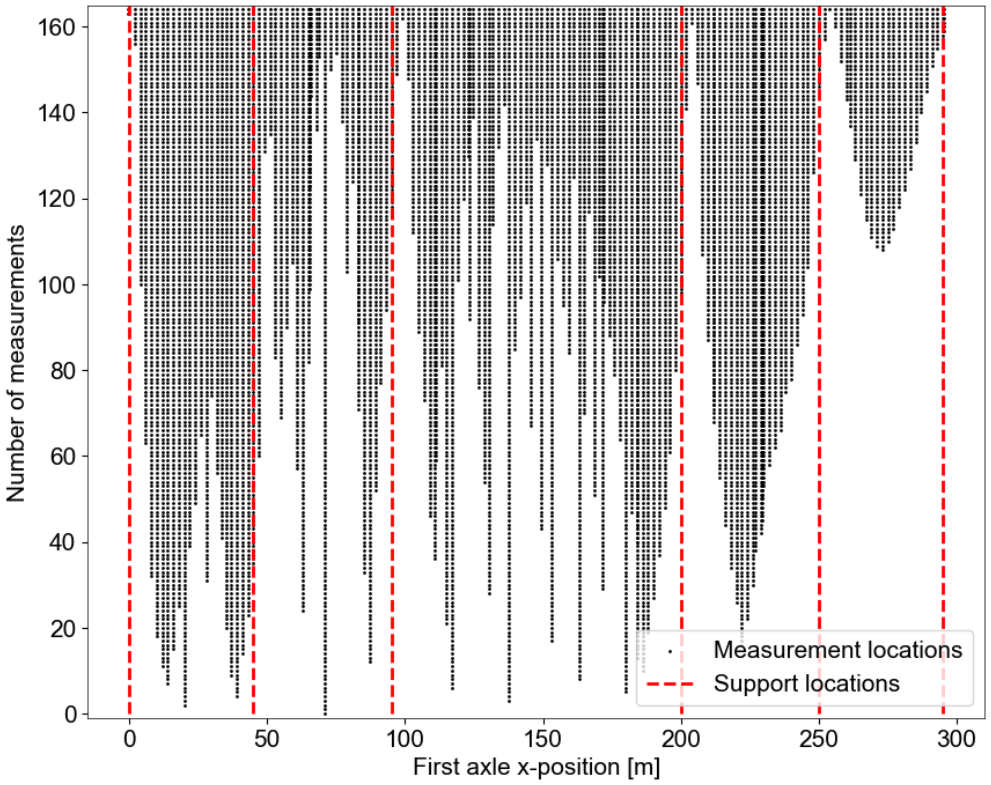


Figure 36: Locations of selected measurements per N.

The coupled probabilistic-physical models are defined as shown in Table 15. The posterior mean and standard deviation of C_v as a function of N is plotted in the figure below. It is noted that the measurements used in the reference model are selected manually as shown in Figure 14 and are not the same as those obtained by measurement selection.

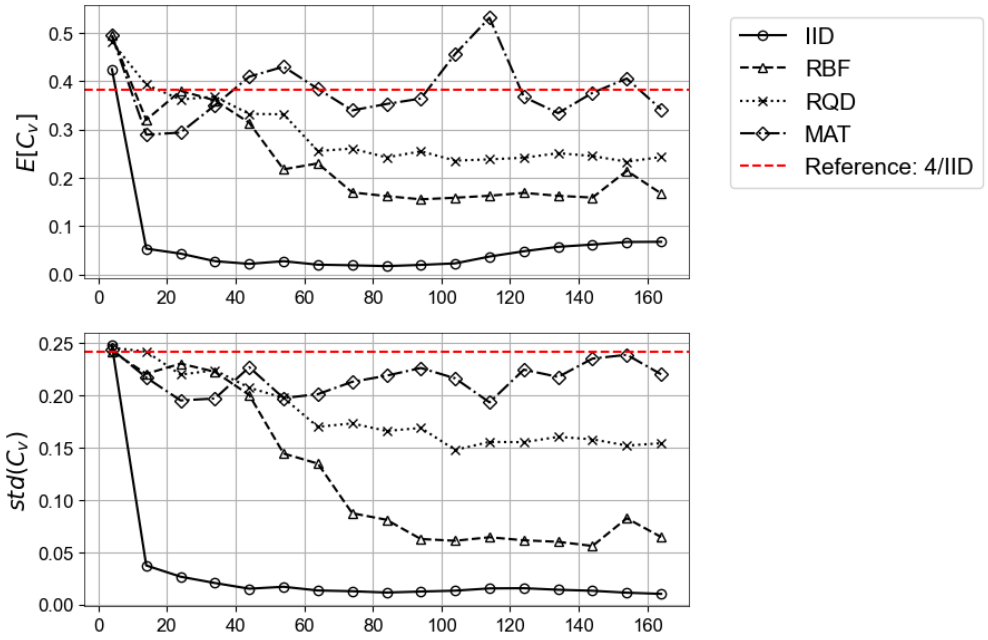


Figure 37: Posterior means and standard deviations of the model prediction uncertainty as a function of N.

It is noted that the damped cosine model is not included in the analysis due to computational issues, likely stemming from the unidentifiability of the parameters. Under the independence assumption, the posterior variance of C_v drastically drops when additional measurements are used up to $N \approx 15$. A more gradual decrease of the uncertainty is observed for the correlated models with the plotted curves becoming flatter up until $N \approx 100$, with the exception of the Matern model. The behaviour of the mean and standard deviation of the inferred C_v as a function of N provide additional indication that it may be necessary to account for dependencies for larger numbers of parameters. This is because based on the inferred correlation lengths shown in Table 16, correlation may affect the inference even for a small number of manually selected measurements. The means and standard deviations of physical model parameters as functions of N are shown in Figure 38.

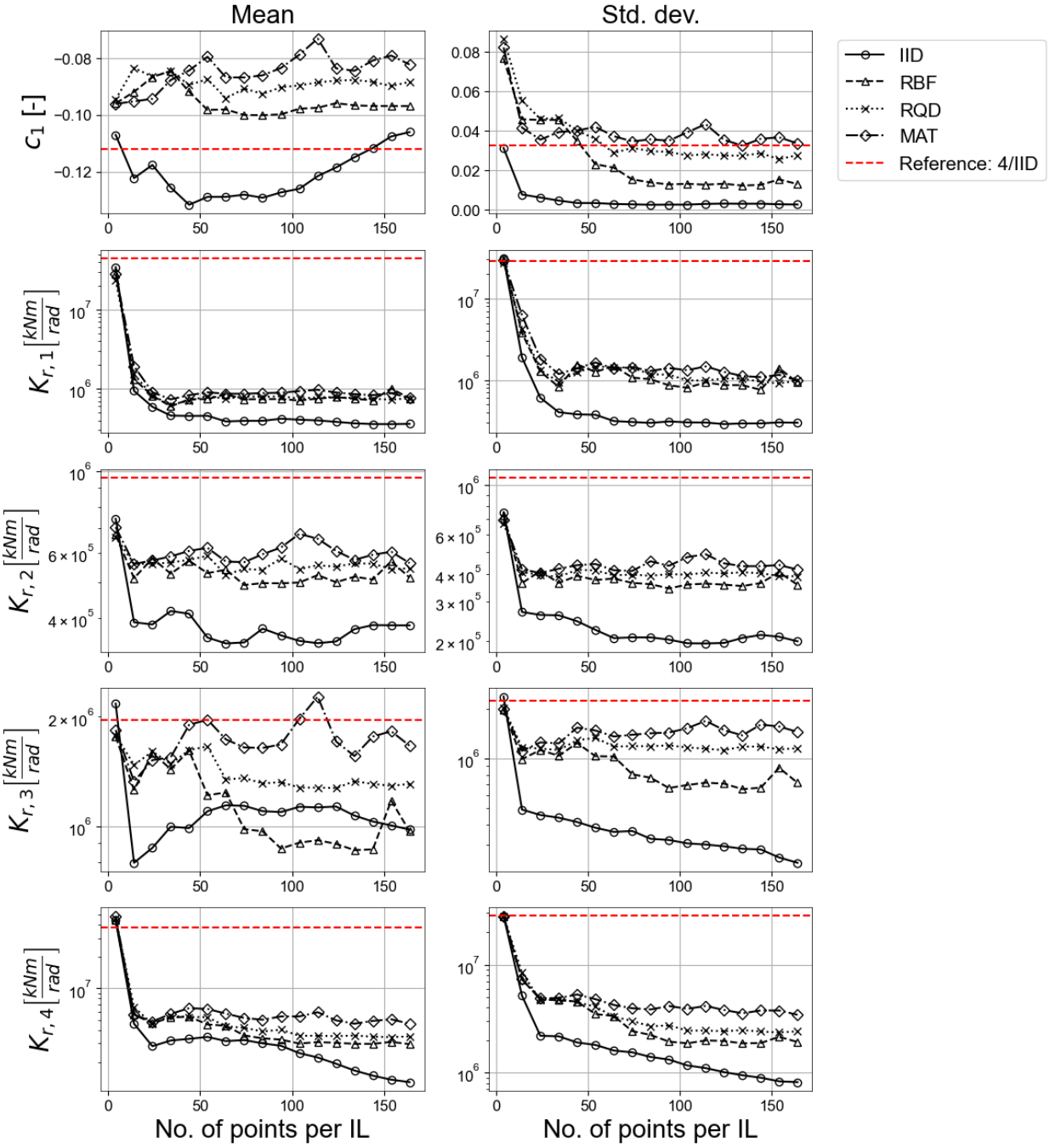


Figure 38: Posterior means and standard deviations of physical model parameters as a function of N.

Observing the standard deviation of the physical model parameters, it can be seen that significantly lower posterior uncertainty is obtained for the IID case for all values of N. In contrast for the models with correlation there is a plateau in the posterior uncertainty for different values of N depending on the parameter. For $K_{\mathrm{r},1}$ and $K_{\mathrm{r},4}$, this occurs at $N = 40$ and $N = 100$ points respectively. Additionally, for these parameters over 15 and 20 points are needed respectively to achieve similar posterior standard deviation with the IID model as

for the RBF model. Based on the inferred correlation lengths, the dependencies in the model prediction uncertainties can not be disregarded in that case. This reinforces the conclusion that the number of parameters that can be inferred under the assumption of independence is limited, while additionally indicating that there are limitations on which parameters can be inferred.

6.4 Case 3: Inference with data from multiple sensors

6.4.1 Impact of considering multiple sensors and combined spatial and temporal dependencies

In this section, influence lines from multiple sensors are used in the inference. An additive model prediction error with combined spatial and temporal correlation is assumed, while the measurement errors at each observation point are taken as i.i.d. Gaussian random variables. The data generating process is described in Section 3.2.1. The discrepancies between measurement and model prediction are modeled as realizations from a Gaussian random field as shown in Section 3.3.2. Three models are used, labeled $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$, distinguished by the size of the dataset and statistical model. Under $\mathcal{M}_1$, complete independence is assumed in the model prediction error and the full influence line for each sensor is used. For $\mathcal{M}_2$ the model uncertainties are assumed to be exponentially correlated in both space and time. A single correlation length parameter l_{corr} is assumed over both the spatial and temporal dimension. The implications of this assumption were not investigated, however, it is supported by the fact that both the spatial and temporal correlation are defined for points along the longitudinal axis of the bridge, and will therefore be affected by the same physical processes. Additional analysis into the validity of this assumption is needed. Finally, for the reference model $\mathcal{M}_3$ four peaks per influence line are used under the assumption of independence. The details of Case 3 are summarized in Table 20.

Table 20: Summary of parameters for case 3.

Case #:	3	Sensors:	H1, H4, H5, H9, H10
Physical model:	Single girder	$\boldsymbol{\theta}_{\text{s}}$:	$c_1, K_{\text{r1}}, K_{\text{r2}}, K_{\text{r3}}, K_{\text{r4}}$
Modeling uncertainty:	Additive	Meas. uncertainty:	Additive, $\sigma_{\text{meas}} = 1.0$ MPa

Model #	Correlation	Dataset size	$\boldsymbol{\theta}_{\text{c}}$
1	Independent	181/IL	σ_{model}
2	Exponential	181/IL	σ_{model}, l_{corr}
3	Independent	4/IL	σ_{model}

A subset of the available measurements is used due to the issues in the measurement processing discussed in 5.2.2. The sensors listed in Table 20 are chosen such that:

- The discrepancies between the physical FE model and the measurements are minimized.

- The structural behaviour is similar in both the right and left lanes.
- The sensors are placed at an adequate distance from the supports where beam theory can be assumed to apply.

Using nested sampling, the posterior distribution for the parameters of each model is obtained. It is noted that in this case, the unidentifiability of σ_{model} and l_{corr} observed when only data from a single sensor was used (see Sections 6.2 and 6.3) seems to have a lesser effect on the posterior. This can be seen by examining the joint posterior distribution of these parameters for the 181/IL,EXP model in Figure 39.

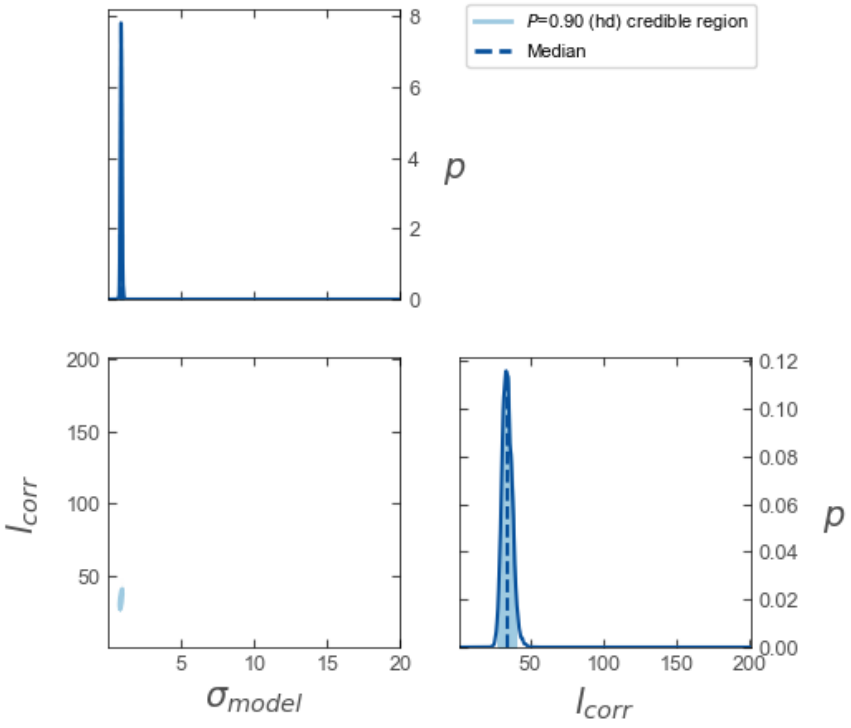


Figure 39: Joint posterior distribution of σ_{model} and l_{corr} for the 181/IL,EXP model.

Furthermore, in the second case (Section 6.3) the lack of data resulted in wide CIs for $K_{\text{r},1}$ and K_{r_4} for the 4/IL,IID reference model. The additional influence lines considered in this case allow for inference of these parameters. To determine the impact of considering multiple sensors and combined spatial and temporal dependencies in the inference of uncertain parameters, the credible intervals of the marginal posteriors are compared in Figure 40. A linear scale is used for all parameters.

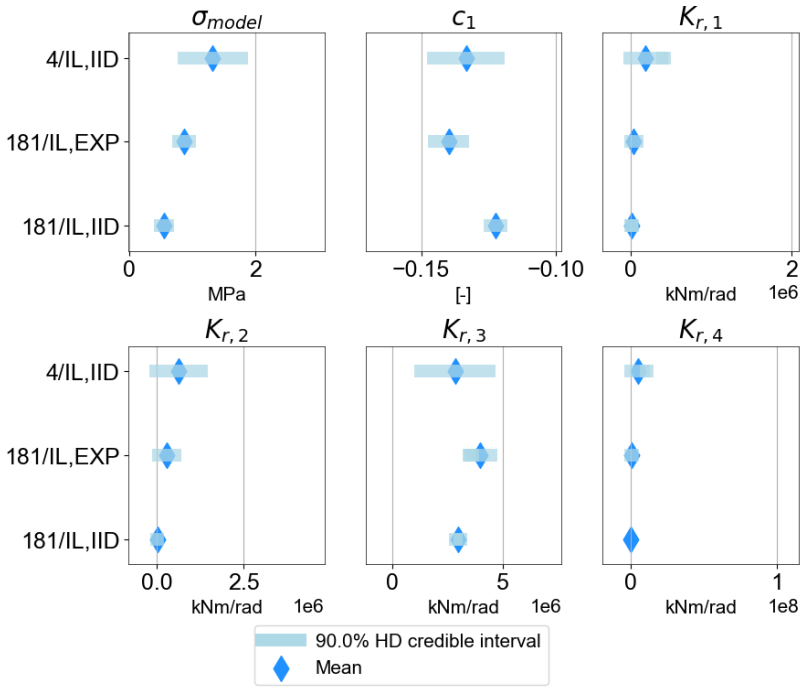
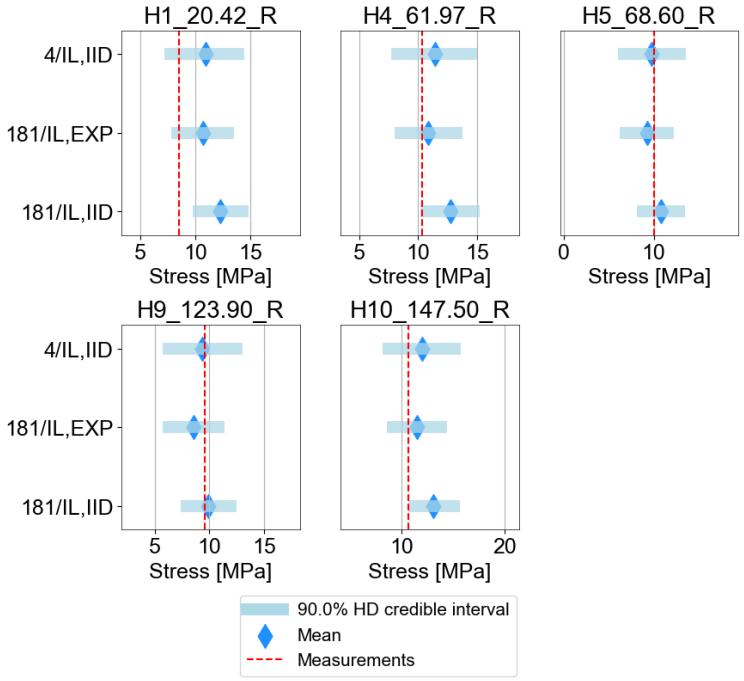


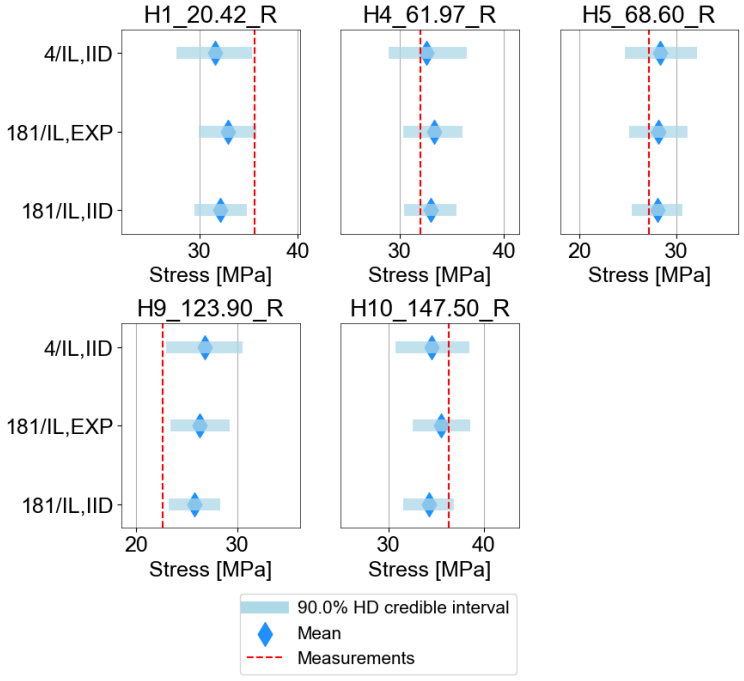
Figure 40: Mean and 90% credible intervals for models $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$.

The additional measurements used in the 181/IL,IID and 181/IL,EXP models result in reduced uncertainty and narrower credible intervals for all parameters, compared to the reference 4/IL,IID model. This is also partially attributed to the fact that no filtering is performed for the measured influence lines. Each measurement may be shifted from an assumed static value (the measured stress if the loading was static) by up to the amplitude of the vibration at that point. This is not expected to have a significant effect for the 181/IL,EXP and 181/IL,IID due to the large number of measurements used, but could affect the 4/IL,IID case substantially.

In order to assess the benefit of considering large datasets while taking into account correlations in the model prediction errors, the quality of the model fit is determined using the posterior predictive stress distribution at the peak of the influence line. This quantity is relevant to reliability analysis, where an accurate prediction of the peak stress at a given location is of importance. The credible intervals and means of the predicted stress distribution at the peak for the different models are compared in Figure 41 for the left and right lane. The sensor for which each influence line is calculated is also denoted, while the dashed red lines indicate the corresponding measured peak stresses.



(a) Left lane.



(b) Right lane.

Figure 41: Comparison of credible intervals of the posterior predictive stress distributions at influence line peaks for each sensor and model.

Given that the used in the comparison were also used in the inference, it is expected that the measured peak stresses will generally be within the plotted 90% credible intervals. This

is confirmed by the results shown in Figure 41, indicating that the standard deviation of the model prediction error σ_{model} is accurately inferred. Furthermore, the lower inferred σ_{model}, as well as the lower uncertainty in the posterior distribution result in narrower credible intervals for the 181/IL,IID and 181/IL,EXP models. To quantify the reduction in the posterior predictive uncertainty, the bounds of the 90% CIs per sensor for each model are provided in Table 21 and Table 22 for the left and right lane respectively.

Table 21: Summary of posterior predictive CIs at left lane influence line peaks (in MPa).

Model	CI	$\bar{y}_{\text{H1,peak}}$	$\bar{y}_{\text{H4,peak}}$	$\bar{y}_{\text{H5,peak}}$	$\bar{y}_{\text{H9,peak}}$	$\bar{y}_{\text{H10,peak}}$
181/IL,IID	Min.	10.35	10.64	8.86	7.83	11.40
	Max.	14.38	14.67	12.81	11.96	15.14
	Range	4.03	4.03	3.95	4.13	3.74
181/IL,EXP	Min.	8.38	8.41	6.92	6.20	9.23
	Max.	13.18	13.19	11.48	10.89	14.08
	Range	4.80	4.78	4.56	4.69	4.85
4/IL,IID	Min.	8.04	8.26	7.02	6.33	9.04
	Max.	14.07	14.51	13.02	12.45	15.19
	Range	6.03	6.25	6.00	6.12	6.15

Table 22: Summary of posterior predictive CIs at right lane influence line peaks (in MPa).

Model	CI	$\bar{y}_{\text{H1,peak}}$	$\bar{y}_{\text{H4,peak}}$	$\bar{y}_{\text{H5,peak}}$	$\bar{y}_{\text{H9,peak}}$	$\bar{y}_{\text{H10,peak}}$
181/IL,IID	Min.	30.09	30.82	26.01	23.66	32.33
	Max.	34.08	34.77	29.97	27.80	36.09
	Range	3.99	3.95	3.96	4.14	3.76
181/IL,EXP	Min.	30.64	30.60	25.87	23.95	33.22
	Max.	35.30	35.37	30.45	28.60	37.85
	Range	4.67	4.78	4.57	4.65	4.63
4/IL,IID	Min.	28.13	29.36	25.19	23.53	31.48
	Max.	34.69	35.91	31.53	29.69	37.90
	Range	6.57	6.55	6.35	6.17	6.42

The calculated 90% posterior predictive CI ranges show a $20\% - 29\%$ reduction for the 181/IL,EXP model compared to the 4/IL,IID model. This indicates that in the case of the IJsselbridge, performing Bayesian inference with large datasets and considering dependence in the model prediction error can yield a benefit in prediction accuracy. This provides motivation for further investigation into the approach in order to assess the benefit in other types of structures and problem domains.

6.5 Concluding summary

The IJsselbridge is used as a real-world case study to assess the feasibility of performing Bayesian system identification for large datasets when dependencies may be present in the model prediction error. The approach proposed in Section 1.3 is implemented for three cases, each with a different pool of candidate models parametrized by the combined set of physical model parameters and statistical model parameters.

In the first case, a single physical model parameter is considered, in order to focus on inferring the statistical model parameters. Uniform priors are placed on all parameters and the posterior is calculated using measurements from a single sensor. The results from models where correlation is included in the probabilistic model formulation and 165 measurements are used per influence line (and also model 165/IL,IID where independence is assumed) are compared with a reference model using four independent measurements per influence line. All parameters are identified with various degrees of posterior uncertainty, with the exception of the correlation length for the damped cosine model where no reduction of the uncertainty is obtained compared to the uniform prior. Subsequent Bayesian model selection indicates that the 165/IL,IID model performs the worst compared to the other candidate models and results in practically zero posterior probability. This initial case demonstrates the feasibility of inferring the structure and parameters of the dependency in Bayesian system identification, however, unidentifiability of the joint posterior of the correlation length and marginal variance parameters may result in wide credible intervals and high point estimates of the model prediction uncertainty. This results in wide uncertainty bands in the posterior predictive stress distribution, e.g. for the Matern model and rational quadratic models, while no significant improvement in model fit or prediction accuracy is observed in comparison to the reference model.

A similar analysis with additional physical model parameters is performed for the second case considering the rotational stiffness parameters of the springs representing the bearings at pillars F, G, H and J (see Figure 12). This case is aimed at studying the effect of the number of inferred parameters and the dataset size on the Bayesian system identification. The results highlight a potential benefit of considering larger datasets and accounting for correlation, as more parameters can be identified in comparison to the reference model. Significant reduction in the credible region for the posterior and the posterior predictive credible intervals are observed. Additionally, the FIM method discussed in Section 3.5.1 is shown to provide a good approximation of the information entropy with minimal computational cost. It is applied for selecting the most informative measurements and studying the effect of the number of measurements on the inference. It is concluded that when a small number of points is used in inference under the assumption of independence in the model prediction uncertainties, not all parameters can be inferred, e.g. the rotational stiffnesses of the springs at supports F and J. Considering dependencies allows for a large number of measurements to be used and makes it possible to infer those parameters.

Finally, in the third case Bayesian system identification is performed using measurements from sensors H1, H4, H5, H9 and H10, assuming an additive model prediction error. The

pool of candidate models is composed of a model with combined spatial and temporal dependencies in the model prediction error, assumed to be described by an exponential kernel, and two models where independence is assumed, differing in the size of the dataset. The efficient likelihood evaluation method detailed in Section 4.3 is utilized to reduce computational time for the model with exponential correlation and make inference feasible for the 5×181 grid of observations. The additional data considered in this case results in a reduction of uncertainty for the correlated model, both in the posterior and posterior predictive distributions, compared to the reference model. This case shows both the feasibility and some potential advantages of the proposed approach for performing system identification for large datasets with spatial and temporal dependencies.

7 Conclusions and future work

In the final section of this report, the conclusions drawn from synthetic examples and the IJssel bridge case study are presented. The research questions posed in Section 1.2 are recapitulated and answered. Also additional insights on the application of Bayesian system identification and sensor placement optimization are provided. This includes practical recommendations for the application of the described approach and an evaluation of the potential benefit compared to the approach applied in the established Bayesian system identification literature. Finally, we indicate some topics where additional research is recommended.

7.1 Conclusions

The research sub-questions posed in Section 1.2 are answered based on the results from the analyses presented throughout this thesis.

1. How can we quantify and model the dependencies in the model prediction error?

To capture the a-priori unknown structure of the dependencies in the model prediction error, a number of candidate probabilistic models is defined. By modeling the discrepancies between the measurements and model prediction as dependent, normally distributed random variables, the correlation can be described in terms of a kernel function and directly incorporated in the formulation of the likelihood function. The Bayesian statistical framework makes it possible to infer the posterior distribution of the parameters of the kernel functions, in addition to physical model parameters, and perform model selection for the pool of candidate models. The analyses presented in Sections 6.2 and 6.3 demonstrate how the parameters of the statistical model can be inferred from the data while the results of the model selection strongly favour the models where correlation is taken into account. For both cases the model where independence is assumed yielded a practically zero posterior probability, indicating that the assumption of correlation provides a better description of the data.

2. How can we efficiently perform system identification using large datasets and computationally demanding likelihood functions and evidence?

Evaluation of the likelihood function for large datasets incurs a significant computational cost. In Section 4, two different approaches based on tridiagonal inverse of the covariance matrix for the Matern kernel family and Kronecker structure of the combined space and time covariance were utilized. For the case of additive model prediction error, the sparse inverse combined with an efficient likelihood evaluation method based on the eigendecomposition of the Kronecker product yields an over 10^4 times faster likelihood evaluation on a grid of 64×64 observations. For the case of multiplicative model prediction error, an efficient method for likelihood evaluation is proposed. This approach achieves a reduction in computational complexity by exploiting the Kronecker structure and tridiagonal inverse of the covariance matrix. A symmetric tridiagonal block Cholesky decomposition can be applied to significantly reduce the complexity of calculating the inverse and determinant of the covariance matrix. For combined spatial and temporal dependencies this method was

shown to be 900 times faster than a naive evaluation on a 64×64 grid, while in the case of correlation in a single dimension the complexity is linear in the number of points, allowing for exact likelihood evaluation for $N > 10^6$ in under a second. Both methods avoid directly forming the full covariance matrix and significantly reduce storage requirements, which can be prohibitive for large problems when the likelihood is naively evaluated. Despite the disadvantage of imposing assumptions on the correlation structure, these approaches lead to significant reduction in computational effort without approximation. To efficiently estimate the evidence when numerical calculation of the high dimensional evidence integral is not feasible, the nested sampling technique is utilized. Through the analyses of the three cases, nested sampling was shown to be an effective approach for sampling the posterior and calculating the evidence for the number of parameters considered (up to 8) with no additional computational cost.

3. How can we efficiently perform sensor placement optimization?

A sensor placement optimization method based on the approximation of the differential entropy using the FIM is implemented for selecting the most informative set of measurements from the given influence lines, and compared with a method based on the calculation of the KL-divergence. The latter requires a full Bayesian inference to be performed for each set of measurements evaluated, which would result in several hours of computational time for the case of the IJsselbridge. In comparison the approximate FIM method requires only the evaluation of the model gradients for each physical parameter at selected nominal values, reducing this time to seconds. Assumptions regarding the correlation structure can be taken into account to determine the optimal set of measurements when dependencies in the model prediction error are known to be present. By comparison to the KL-divergence method, the accuracy of the approximation is deemed to be sufficient for application to the case of the IJsselbridge.

4. What is the impact of considering dependencies in the model prediction error in Bayesian system identification?

It was shown in the analyses presented in Sections 6.2 and 6.3 that considering correlations by inferring the uncertainty parameters from the data resulted in a reduction of the effective sample size and wider posterior CIs. In these cases the unidentifiability of the marginal variance and correlation length parameters resulted in high estimates of the modeling uncertainty. This effect is highly dependent on the probabilistic model, with the Matern and Rational quadratic models obtaining similar or larger posterior and posterior predictive uncertainty than the reference case, as quantified by the calculated credible intervals. For the Radial basis function model low uncertainty in the posterior and posterior predictive was obtained, while for the Damped cosine model the correlation length parameter was not identified. Additionally, in the second case (see Section 6.3) taking into account the dependencies allowed for 165 points per influence lines to be used, resulting in accurate inference of the full set of physical model parameters, which was not achieved when only the influence line peaks were used. Similarly, for the case of inference using data from multiple sensors (see

6.4) lower uncertainty in the posterior and more accurate posterior predictive distribution of the peak stress were obtained when the full dataset was considered.

Through the answers to the sub-questions listed previously, the main research question is addressed:

How to perform Bayesian system identification on real-world civil and offshore structures in a practically feasible/manageable way, when spatial and/or temporal dependence might be present?

Through the IJsselbridge case study, it is demonstrated how Bayesian system identification can be feasibly performed when spatial and temporal dependence might be present. Datasets with up to 2×900 measurements are used to infer the uncertain parameters of both the statistical and physical model parameters, and Bayesian model selection is applied to determine the most suitable probabilistic model. In addition to updating the physical model and quantifying the uncertainty in the physical model parameters, this approach makes it possible to infer the correlation structure and quantify uncertainties in the model prediction. The Matern, Radial basis and Rational quadratic kernel functions were found to provide a good approximation of the correlation structure. To make the approach feasible for the large datasets (e.g. $N > 10^3$) in the case of temporally and spatiotemporally correlated multiplicative model prediction error with additive i.i.d. Gaussian white noise, an efficient method for log-likelihood evaluation is proposed.

Through a real-world case study the implemented approach is compared with a reference case, representing the typical approach followed in the literature. The most important conclusions obtained from the case study are reiterated here:

- The number of measurements used in the system identification was found to have a significant effect on the results. For small numbers of parameters ($N_\theta \leq 3$), sufficient reduction of the posterior uncertainty and a high degree of model fit can be achieved even with four measurements per influence line. Further increase of N under the assumption of correlation did not provide any reduction of uncertainty in the posterior and posterior predictive.
- For the number of physical model parameters necessary to improve model fit, quantify the uncertainties and fully capture the structural behaviour ($N_{\theta_s} = 5$ and $N_{\theta_c} \geq 1$ in Case 2) using four peaks per influence line is not sufficient.
- The use of a few selected peaks under the i.i.d. assumption for the model prediction error in Bayesian inference can lead to insufficient data and inability to infer all of the parameters of interest, therefore limiting the number and type of parameters that can be identified. Using the full dataset under the assumption of dependence allows for additional parameters to be inferred.
- The unidentifiability of the σ_{model} and l_{corr} parameters results in wide a wide CIs for the posterior of σ_{model} and CIs for the posterior predictive distribution of the stress influence line for the Matern and Rational quadratic kernels.

- For additive model prediction uncertainty using data from multiple sensors and considering spatial and temporal dependencies a $20\% - 29\%$ reduction in the posterior predictive CI was obtained at the influence line peaks, compared to a reference model where independence is assumed.

7.2 Future work

This section aims to identify topics where additional research is needed, questions that were not addressed and methods that could potentially improve the feasibility and applicability of the described approach. As discussed in Section 1.4, the scope of this thesis is limited to the description of a practically feasible approach for considering dependencies in the model prediction error in Bayesian system identification, and few conclusions are made on the benefit of this method. Despite this, the results described in Sections 6.3 and 6.4 indicate that there are advantages to this approach compared to the methods typically applied in the Bayesian system identification literature and provide a motivation for further research. In the following, a number of additional research questions are proposed.

7.2.1 Benefits and impact of the approach in practical applications

To the author's knowledge, there are no other applications of this approach to real-world structures in the Bayesian system identification literature. Although it is expected that the conclusions drawn from the IJsselbridge case study could be generalized to a number of different structure types or problem domains, this should be carefully examined. Furthermore, the impact and potential benefits of this approach are not extended to practical engineering applications, such as the prediction of remaining service life, the assessment of structural condition and damage identification, as well as reliability calculations. Investigation of these topics could provide insight into the effect of dependencies in the model prediction error and determine the benefits of considering them in Bayesian system identification.

- Extension of the approach to other types of structures.
- Investigation of the effect of spatial and temporal dependencies on reliability calculations, damage identification and remaining service life prediction.
- Evaluation of prediction accuracy for locations in the structure not used in inference.

7.2.2 Additional research into efficient evaluation of the likelihood

The proposed efficient likelihood evaluation method provides a substantial reduction in computational complexity, but is limited to a single family of kernels. Furthermore, this method can not account for correlation in more than two dimensions. It is expected that these issues can be addressed by utilizing more sophisticated approaches that can be generalized to other correlation structures, better describe the correlation in the model prediction error and potentially take into account the measurements as a continuous record instead of discrete points. These points are summarized below:

- How can other efficient log-likelihood evaluation methods such as state space models, particle filtering and approximate methods be utilized in Bayesian system identification for structures?
- How can the continuous record log-likelihood be taken into account?
- How can the proposed efficient likelihood approach be extended to correlations in more than two dimensions?

7.2.3 Additional consideration on sensor placement optimization

Sensor placement optimization was only briefly touched upon in this thesis. The quality of the approximation of the FIM approach was evaluated for the case of the IJsselbridge and no general conclusions on its applicability were drawn. Additionally, the application was limited to selecting a set of informative measurements and not extended to sensor layout optimization for point-like strain gauges or line-like fiber optic strain gauges. A summary of additional research questions on the topic is provided below:

- How can the FIM approximation method be applied for efficient sensor layout optimization?
- How can sensor layout optimization be applied under combined spatial and temporal correlations?
- How can sensor placement optimization be performed for the case of linear fiber optic strain gauges?

7.2.4 Impact of the statistical model assumptions

A number of assumptions regarding the structure of the correlation, the prior distribution of the statistical parameters, as well as the distribution of the model prediction error and the form of the model prediction uncertainties (e.g. additive or multiplicative) are made throughout this thesis. These assumptions are not sufficiently investigated and additional research is needed to gain better insight into how their effect on the results of Bayesian system identification. Answering the following questions could provide insight into the validity and impact of these assumptions.

- How can more suitable priors for the statistical model parameters be defined?
- What is the impact of considering multiplicative modeling uncertainty vs. additive modeling uncertainty?
- What is the impact of the separable space-time covariance assumption?
- What is the impact of using other (i.e. not Gaussian) distributions to describe the model prediction and measurement error?
- What is the influence of different kernel functions on the posterior distributions?

- How are the results of model selection affected by the different kernel functions?
- Is there a benefit to considering complicated kernel functions that can more accurately describe the correlation structure?

References

Abdulah, S., Ltaief, H., Sun, Y., Genton, M. G., and Keyes, D. E. (2018). ExaGeoStat: A High Performance Unified Software for Geostatistics on Manycore Systems. *IEEE Transactions on Parallel and Distributed Systems*, 29(12):2771–2784.

Ang, A. and Tang, W. H. (2007). *Probability Concepts in Engineering Emphasis on Applications to Civil and Environmental Engineering*. Wiley.

Astroza, R., Ebrahimian, H., Li, Y., and Conte, J. P. (2017). Bayesian nonlinear structural FE model and seismic input identification for damage assessment of civil structures. *Mechanical Systems and Signal Processing*, 93:661–687.

Bar-On, I., Codenotti, B., and Leoncini, M. (2006). A Fast Parallel Cholesky Decomposition Algorithm for Tridiagonal Symmetric Matrices. *SIAM Journal on Matrix Analysis and Applications*, 18(2):403–418.

Barrias, A., Casas, J. R., and Villalba, S. (2016). A Review of Distributed Optical Fiber Sensors for Civil Engineering Applications. *Sensors, 16(5), 748.*

Bathe, K.-J. (2006). *Finite element procedures*. Klaus-Jurgen Bathe.

Beck, J. and Katafygiotis, L. (1998). Updating Models and Their Uncertainties. i: Bayesian Statistical Framework. *Journal of Engineering Mechanics. 124:455-461.*

Beck, J. and Yuen, K. (2004). Model Selection Using Response Measurements: Bayesian Probabilistic Approach. *Journal of Engineering Mechanics*, 130(2):192–203.

Beck, J. L. (2010). Bayesian system identification based on probability logic. *Struct. Control and Health Monitoring 17:825 - 847.*

Behmanesh, I. and Moaveni, B. (2014). Probabilistic identification of simulated damage on the Dowling Hall footbridge through Bayesian finite element model updating. *Structural Control and Health Monitoring*, 22(3):463–483.

Brownjohn, J. (2007). Structural health monitoring of civil infrastructure. *Phil. Trans. R. Soc. A*, 365:589–622.

Cervenka, V., Cervenka, J., and Kadlec, L. (2018). Model uncertainties in numerical simulations of reinforced concrete structures. *Structural Concrete*, 19(6):2004–2016.

Chai, X. (2019). Probabilistic system identification and reliability updating for hydraulic structures - Application to sheet pile walls. *MSc. thesis.*

Chen, H.-P. (2018). *Structural Health Monitoring of Large Civil Engineering Structures*. Wiley-Blackwell.

Cheong, S. (2016). *Parameter Estimation for the Spatial Ornstein-Uhlenbeck Process with Missing Observations*. PhD thesis, University of Wisconsin Milwaukee UWM.

Chiachío, J., Chiachío, M., Saxena, A., Sankararaman, S., Rus, G., and Goebel, K. (2015). Bayesian model selection and parameter estimation for fatigue damage progression models in composites. *International Journal of Fatigue*, 70:361–373.

Coakley, E. S. and Rokhlin, V. (2013). A fast divide-and-conquer algorithm for computing the spectra of real symmetric tridiagonal matrices. *Applied and Computational Harmonic Analysis*, 34(3):379–414.

Diggle, P. J. and Ribeiro, P. J. (2002). Bayesian Inference in Gaussian Model-based Geostatistics. *Geographical and Environmental Modelling*, 6(2):129–146.

Duvenaud, D. K. (2014). *Automatic Model Construction with Gaussian Processes*. PhD thesis, University of Cambridge.

Ebrahimian, H., Astroza, R., Conte, J. P., and Papadimitriou, C. (2018). Bayesian optimal estimation for output-only nonlinear system and damage identification of civil structures. *Structural Control and Health Monitoring*, 25(4).

Fuglstad, G.-A., Lindgren, F., Simpson, D., and Rue, H. (2014). Exploring a New Class of Non-stationary Spatial Gaussian Random Fields with Varying Local Anisotropy. *Statistica Sinica*.

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. (2015). Interpretable Priors for Hyperparameters for Gaussian Random Fields. *arXiv:1503.00256*.

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. (2018). Constructing Priors that Penalize the Complexity of Gaussian Random Fields. *Journal of the American Statistical Association*, 114(525):445–452.

Furrer, R., Genton, M. G., and Nychka, D. (2012). Covariance Tapering for Interpolation of Large Spatial Datasets. *Journal of Computational and Graphical Statistics, 15:3, 502-523*.

Gardner, J. R., Pleiss, G., Bindel, D., Weinberger, K. Q., and Wilson, A. G. (2018). GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration. *NeurIPS 2018*.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis, Third edition*. Chapman and Hall/CRC.

Ghali, A. and Neville, A. (2017). *Structural Analysis : A Unified Classical and Matrix Approach, Seventh Edition*. Taylor and Francis.

Goller, B. and Schueller, G. I. (2011). Investigation of model uncertainties in Bayesian structural model updating. *Journal of Sound and Vibration*, 330(25-15): 6122–6136.

Green, P. L., Cross, E. J., and Worden, K. (2015). Bayesian system identification of dynamical systems using highly informative training data. *Mechanical Systems and Signal Processing*, 56-57:109–122.

Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1):97–109.

Higson, E., Handley, W., Hobson, M., and Lasenby, A. (2018). Dynamic nested sampling: an improved algorithm for parameter estimation and evidence calculation. *Statistics and Computing 29, 5 (2019) p891-913.*

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science, Vol. 14, No. 4 (Nov., 1999), pp. 382-401.*

Hu, X., Lindgren, F., Simpson, D., and Rue, H. (2013). Multivariate Gaussian Random Fields with Oscillating Covariance Functions using Systems of Stochastic Partial Differential Equations. *arXiv:1307.1384.*

Huang, Y., Shao, C., Wu, B., Beck, J. L., and Li, H. (2019). State of the art review on Bayesian inference in structural system identification and damage assessment. *Advances in Structural Engineering*, 22(6):1329–1351.

Jeffreys, H. (2003). *Theory of probability, 3rd Edition.* Oxford Classic Texts in the Physical Sciences. Clarendon Press, 3 edition.

Katzfuss, M. and Guinness, J. (2019). A General Framework for Vecchia Approximations of Gaussian Processes. *Statistical Science*, 36(1).

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *J. R. Statist. Soc. B, 63, Part 3, pp. 425 - 464.*

Kurz, J. H., Boller, C., and Dobmann, G. (2013). Condition Assessment of Civil Infrastructure in Europe: Recent Developments and What Might Be Ahead. *Journal of Engineering Mechanics*, 139(6):702–711.

Lam, H. F., Hu, Q., and Wong, M. T. (2014). The Bayesian methodology for the detection of railway ballast damage under a concrete sleeper. *Engineering Structures*, 81:289–301.

Litvinenko, A., Sun, Y., Genton, M. G., and Keyes, D. E. (2019). Likelihood approximation with hierarchical matrices for large spatial datasets. *Computational Statistics & Data Analysis*, 137:115–132.

Lye, A., Cicirello, A., and Patelli, E. (2021). Sampling methods for solving Bayesian model updating problems: A tutorial. *Mechanical Systems and Signal Processing*, 159:107760.

MacKay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms.* Cambridge University Press.

Marcotte, D. and Allard, D. (2018). Gibbs sampling on large lattice with GMRF. *Computers & Geosciences*, 111:190–199.

Martinez-Luengo, M., Kolios, A., and Wang, L. (2016). Structural health monitoring of offshore wind turbines: A review through the Statistical Pattern Recognition Paradigm. *Renewable and Sustainable Energy Reviews*, 64:91–105.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

Meurant, G. (1992). A Review on the Inverse of Symmetric Tridiagonal and Block Tridiagonal Matrices. *SIAM Journal on Matrix Analysis and Applications*, 13(3):707–728.

Mthembu, L., Marwala, T., Friswell, M. I., and Adhikari, S. (2011). Model selection in finite element model updating using the Bayesian evidence statistic. *Mechanical Systems and Signal Processing*, 25:2399–2412.

Papadimitriou, C., Beck, J. L., and Au, S.-K. (2000). Entropy-Based Optimal Sensor Location for Structural Model Updating. *Journal of Vibration and Control 6: 781-800.*

Papadimitriou, C. and Lombaert, G. (2012). The effect of prediction error correlation on optimal sensor placement in structural dynamics. *Mechanical Systems and Signal Processing 28 105-127.*

Pasquier, P. and Marcotte, D. (2020). Robust identification of volumetric heat capacity and analysis of thermal response tests by Bayesian inference with correlated residuals. *Applied Energy*, 261:114394.

Pasquier, R. and Smith, I. F. C. (2015). Robust system identification and model predictions in the presence of systematic uncertainty. *Advanced Engineering Informatics*, 29(4):1096–1109.

Quarteroni, A., Sacco, R., and Saleri, F. (2007). *Numerical Mathematics*. Springer.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*, volume . www.GaussianProcess.org/gpml. the MIT Press.

Rogers, T. J. (2018). *Towards Bayesian system identification: With application to SHM of offshore structures*. PhD thesis, University of Sheffield.

Rozsas, A., Allaix, D. L., and Slobbe, A. (2020). Probabilistic system and load identification for bridge 705 using on-site measurements and finite element analysis. Technical report, Internal TNO report.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman Hall/CRC.

Sarkka, S. and Solin, A. (2019). *Applied Stochastic Differential Equations*. Cambridge University Press.

Simoen, E., Papadimitriou, C., De Roeck, G., and Lombaert, G. (1998). *Influence of the prediction error correlation model on Bayesian FE model updating results*. Life-Cycle and Sustainability of Civil Infrastructure Systems.

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

Simoen, E., Roeck, G. D., and Lombaert, G. (2015). Dealing with uncertainty in model updating for damage assessment: A review. *Mechanical Systems and Signal Processing*, 56-57:123–149.

Simpson, D., Lindgren, F., and Rue, H. (2012). Think continuous: Markovian Gaussian models in spatial statistics. *Spatial Statistics*, 1:16–29.

Skilling, J. (2006). Nested Sampling for General Bayesian Computation. *Bayesian Analysis 1, Number 4, pp. 833–860.*

Solin, A. (2016). *Stochastic and Differential and Equation Methods and for and Spatio-Temporal Gaussian and Process Regression.* PhD thesis, Aalto University, Department of Neuroscience and Biomedical Engineering / Department of Computer Science.

Speagle, J. S. (2019). dynesty: A Dynamic Nested Sampling Package for Estimating Bayesian Posteriors and Evidences. *arXiv:1904.02180v1.*

Stegle, O., Lippert, C., Mooij, J. M., Lawrence, N. D., and Borgwardt, K. M. (2011). Efficient inference in matrix-variate Gaussian models with iid observation noise. *Advances in Neural Information Processing Systems 24.*

Sykorka, M., Krejsa, J., Mlcoch, J., Prieto, M., and Tanner, P. (2018). Uncertainty in shear resistance models of reinforced concrete beams according to fib MC2010. *Structural Concrete 1-12.*

Varin, C. (2008). On composite marginal likelihoods. *AStA Advances in Statistical Analysis*, 92(1):1–28.

Vecchia, A. V. (1988). Estimation and Model and Identification for Continuous and Spatial Processes. *J. R. Statist. Soc. B SO, No.2, pp*, 2:297–312.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

Wang, K. A., Pleiss, G., Gardner, J. R., Tyree, S., Weinberger, K. Q., and Wilson, A. G. (2019). Exact Gaussian Processes on a Million Data Points. *arXiv:1903.08114.*

Ye, X. W., Su, Y. H., and Han, J. P. (2014). Structural Health Monitoring of Civil Infrastructure Using Optical Fiber Sensing Technology: A Comprehensive Review. *The Scientific World Journal*, 2014:1–11.

A Appendix A - Model

Additional details on the model are provided in this appendix.

A.1 Sensitivity analysis

The sensitivity of the model response to variations of the physical parameters is examined. The following show the influence lines obtained for a range of values for each of the structural parameters c_1, K_{r1}, K_{r2}, K_{r3} and K_{r4}. Influence lines obtained using the single girder model for sensor PBR-P4 are shown. During the variation of each parameter the following constant values are assigned to all other parameters:

- c_1 = -0.175
- $K_{ri} = 0.0$

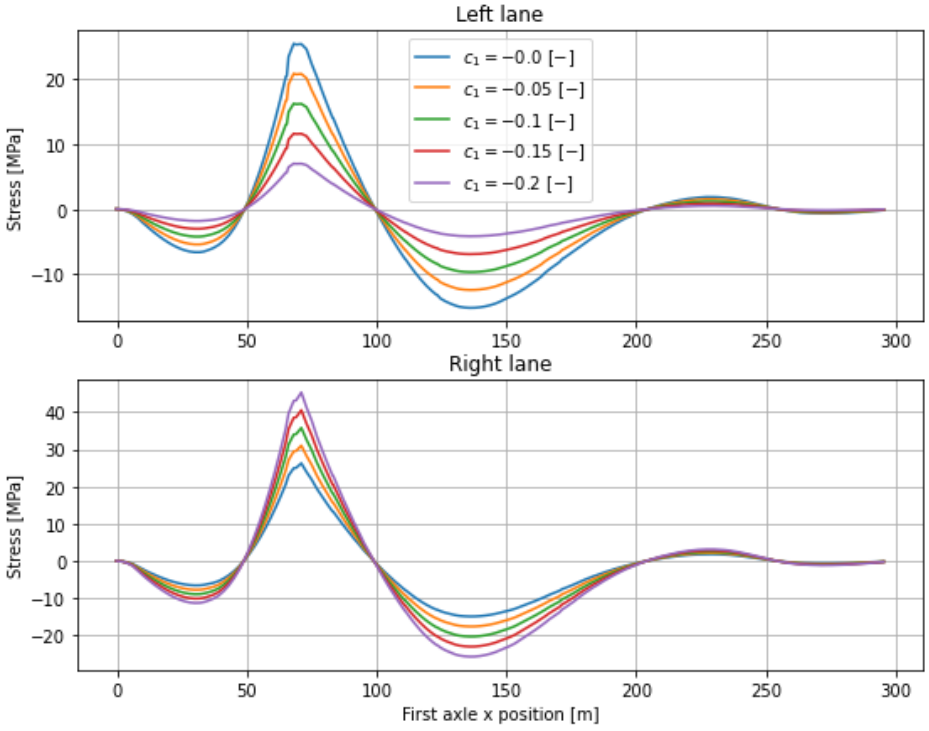


Figure 42: Sensitivity of model response to c_1.

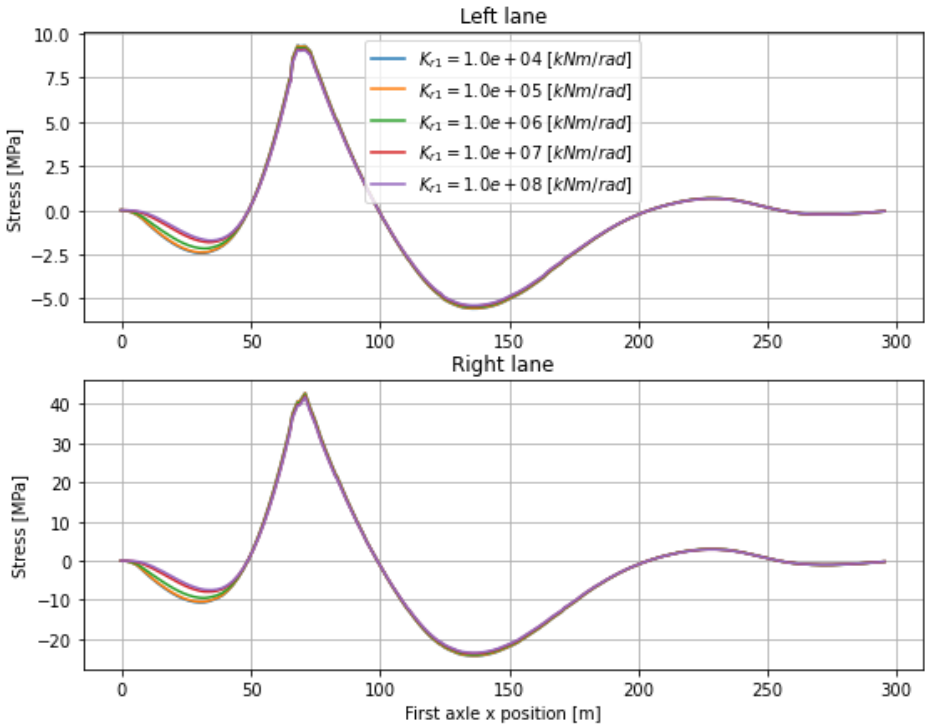


Figure 43: Sensitivity of model response to K_{r1}.

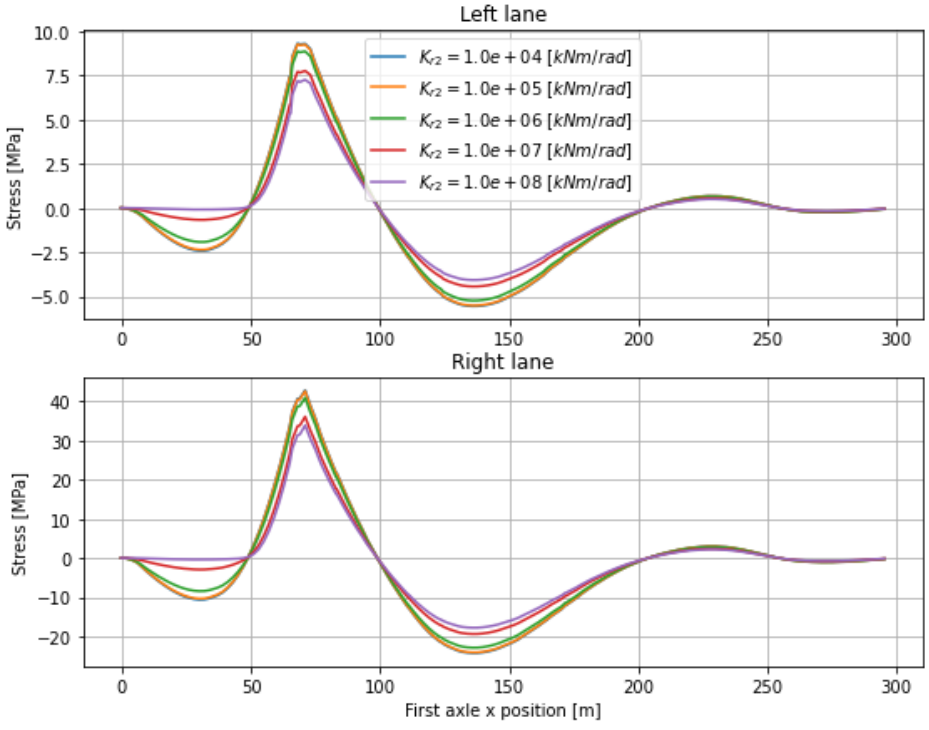


Figure 44: Sensitivity of model response to K_{r2}.

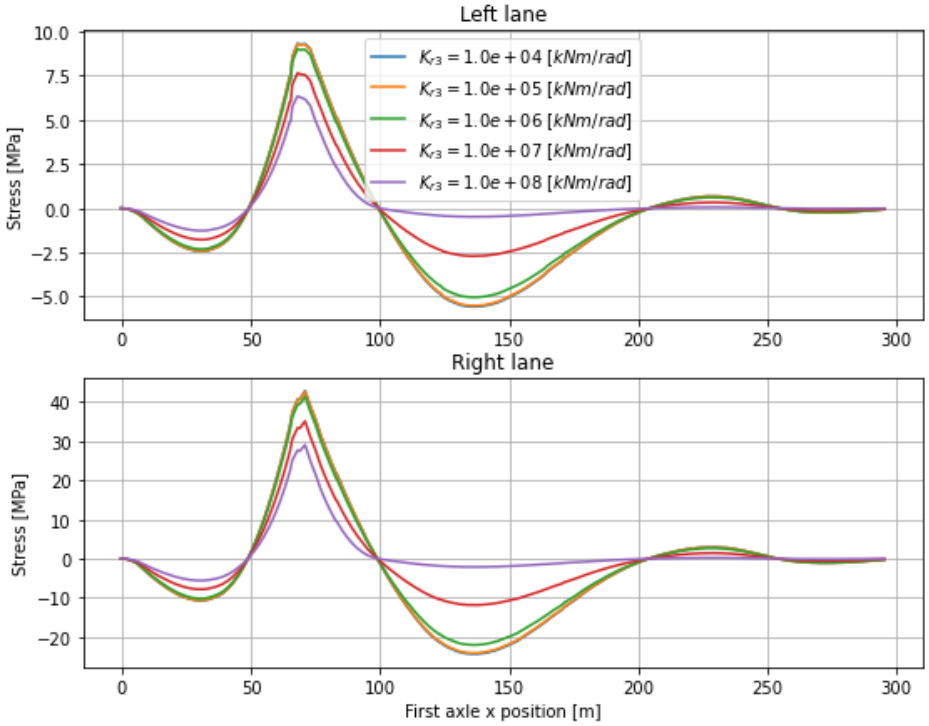


Figure 45: Sensitivity of model response to K_{r3}.

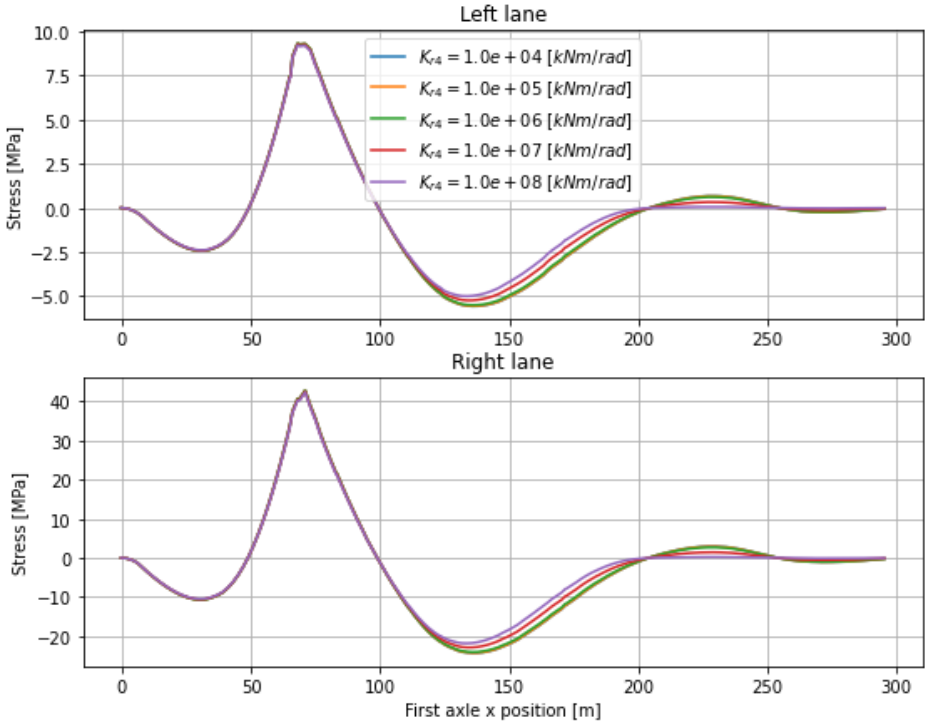


Figure 46: Sensitivity of model response to K_{r4}.

The stress at each sensor for the truck front axle positioned at the sensor location is plotted below as a function of the parameters listed previously.

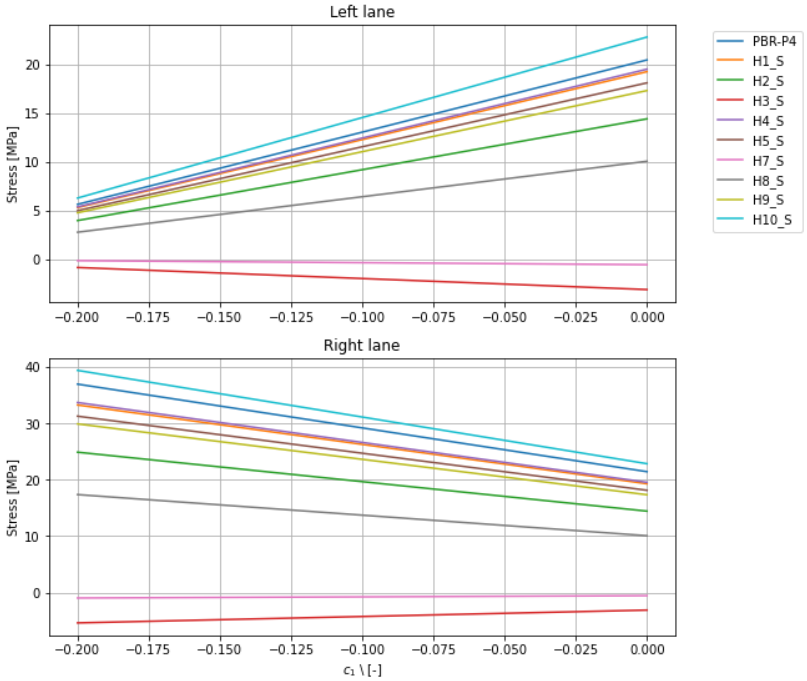


Figure 47: Sensitivity of model response to c_1.

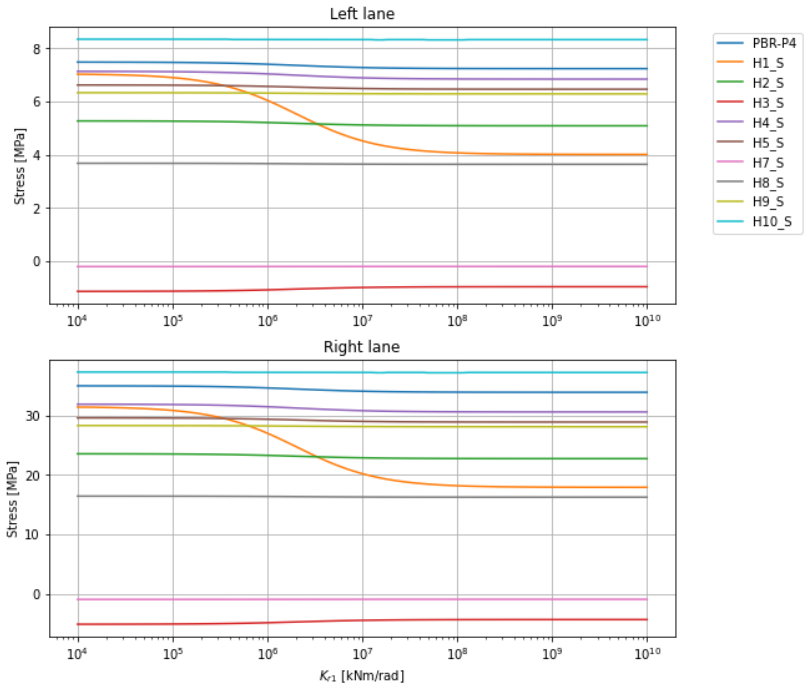


Figure 48: Sensitivity of model response to K_{r1}.

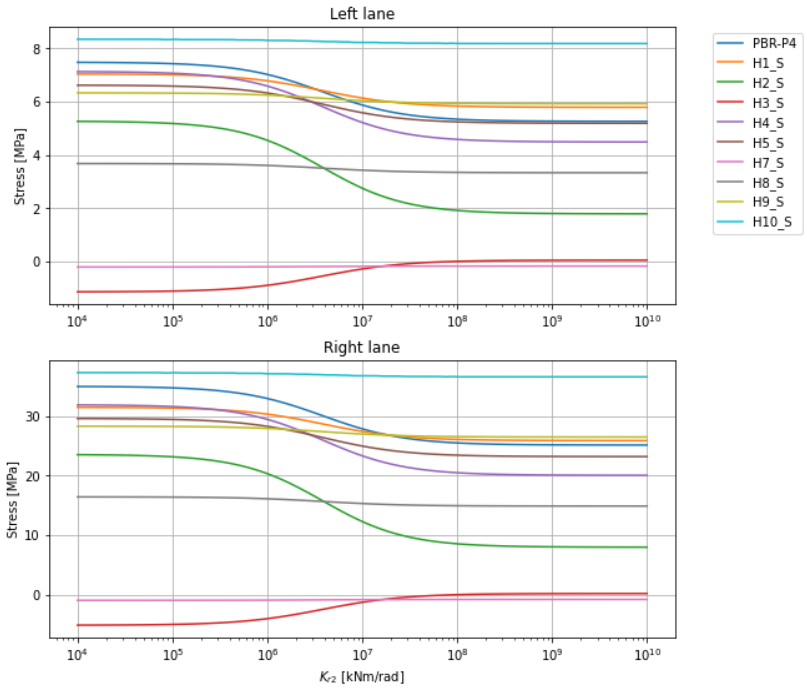


Figure 49: Sensitivity of model response to K_{r2}.

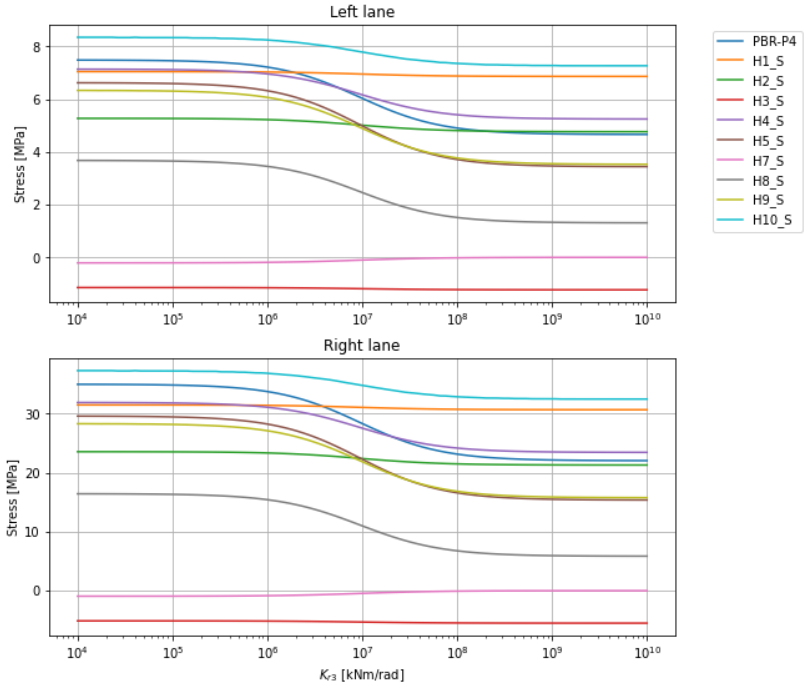


Figure 50: Sensitivity of model response to K_{r3}.

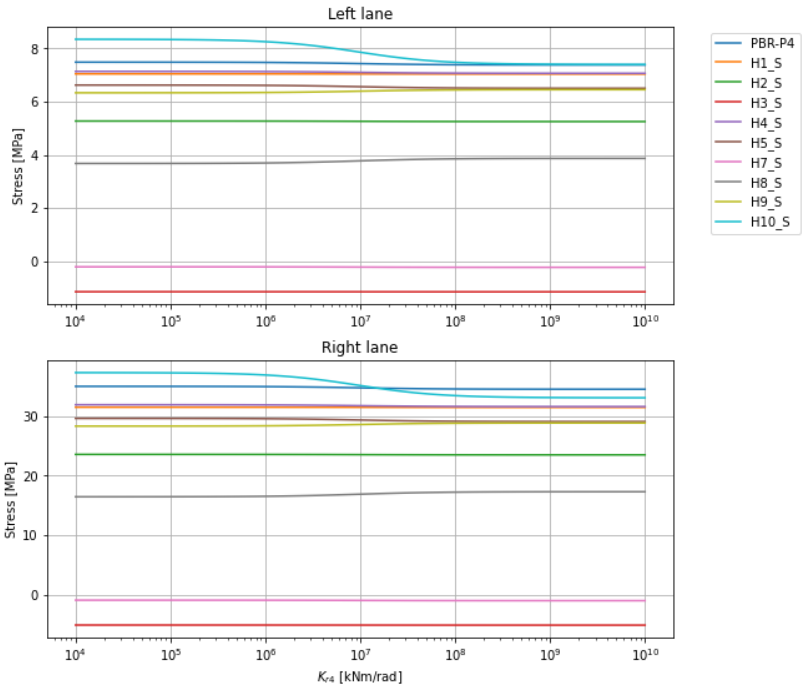


Figure 51: Sensitivity of model response to K_{r4}.

B Appendix B - Measurements

Additional information on the data acquisition, measurement processing, controlled loading test parameters and assumptions are provided in this appendix.

B.1 Measured influence line plots

All influence lines for the Fugro sensors used in this thesis are provided in the figure below. The dashed lines represent the manually selected peaks used in inference for the reference 4/IL,IID case:

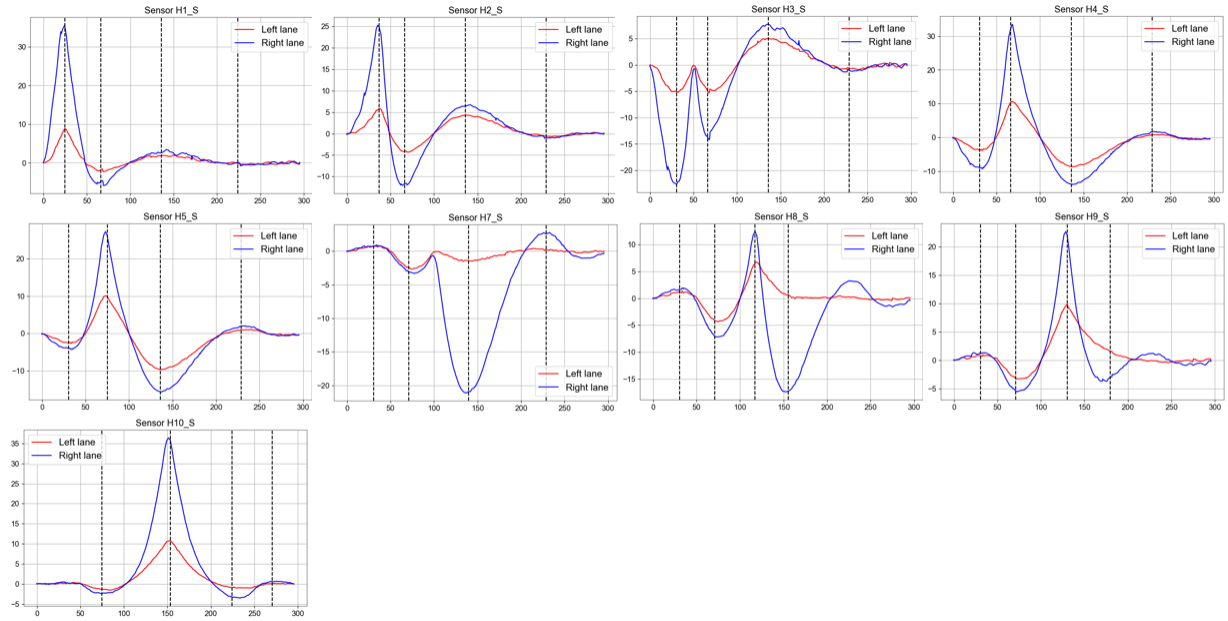


Figure 52: Fugro measurements for all sensors with vertical lines corresponding to the measurements used in the 4/IL,IID case.

B.2 Data acquisition and processing

Measurements are obtained by 34 sensors connected by a total of 9 fiber optic lines to an interrogator sampling at a 50.0 Hz frequency. A total of six tests are performed with trucks driving over the left or right lane at a constant speed, with the truck transverse position roughly corresponding to that of the right or left girder depending on the test. Both the transverse position and speed were manually controlled. A summary of these tests is provided below.

Table 23: Controlled loading test parameters.

Time start [CET]	Lane	Speed [km/h]
21:56:55	Right	20
22:05:55	Left	20
22:21:30	Left	80
22:29:12	Left	80
22:41:25	Right	80
22:49:15	Right	80

The truck center of mass is calculated by assuming that the front axle take 12% of the total load, with the remaining axles taking 22% of the load. The center of mass is calculated as:

$$x_{\text{CM}} = \frac{\sum w_i \cdot x_i}{\sum w_i} \tag{65}$$

During processing it was found that the truck speed deviated from the assumed 20 km/h and this deviation should be accounted for in the processing. To implement the correction it was assumed that the influence line peak for each sensor occurs when the truck center of mass coincides with the sensor longitudinal position. The time difference Δt between the peaks of sensors H1 and H10 was measured. The distance Δx between the two sensor positions was then divided by Δt to obtain the truck velocity for the left and right lanes equal to $v_l = 21.18$ km/h and $v_r = 21.66$ km/h respectively.

The influence lines are obtained by applying a time window to the strain timeseries. The window start and end times correspond to the first track axle entering the bridge and the last truck axle leaving the bridge respectively, as shown in Figure 53. The time corresponding to the start and end position can be determined using the known distances Δx_1 and Δx_2 and the truck speed calculated previously. A -0.1 s shift was applied to the right lane measurements to minimize the discrepancies between the measured and predicted stress influence lines discussed in the following section.

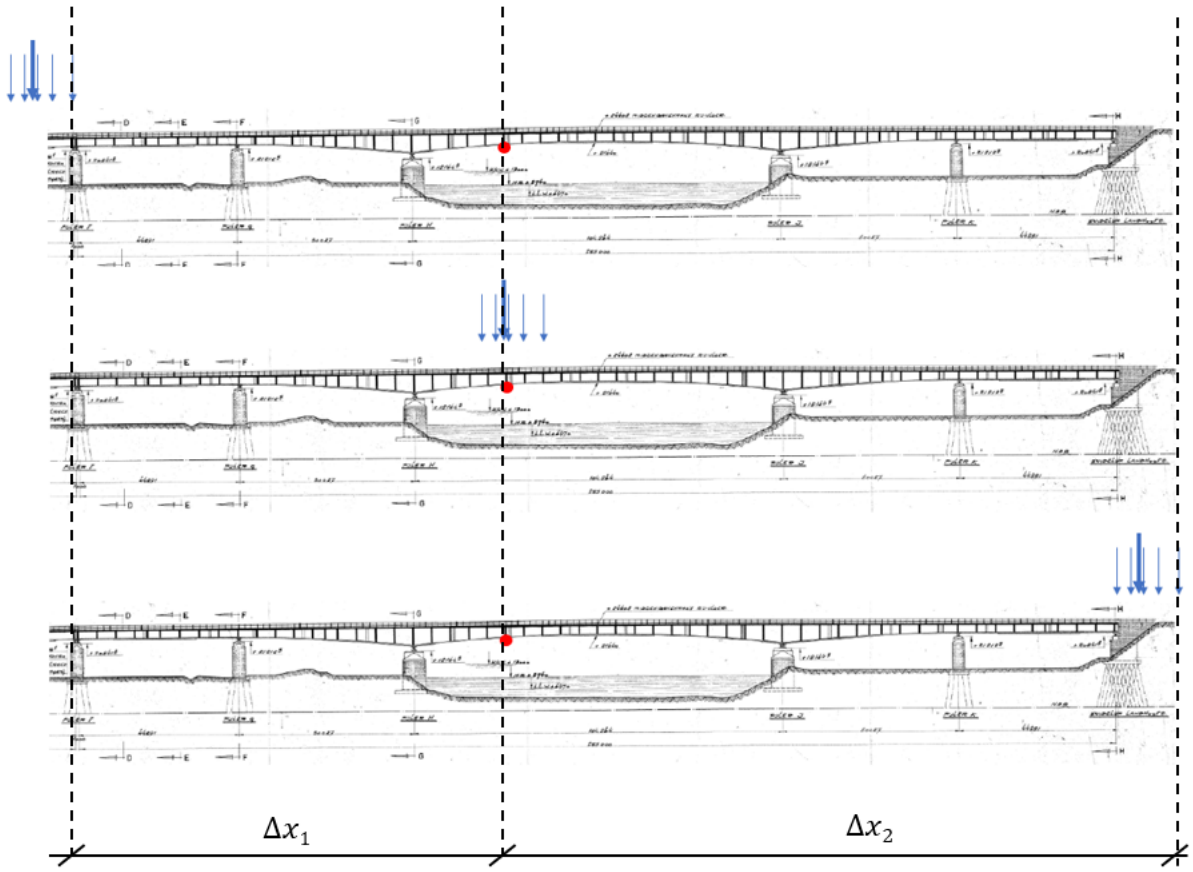
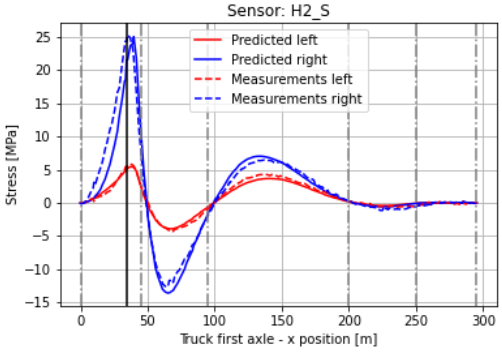


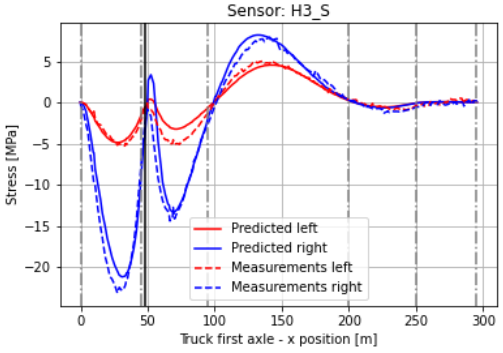
Figure 53: Load position at influence line start (top), peak (middle) and end (bottom).

B.3 Measurement and model discrepancies

Discrepancies between the measured and predicted stress influence lines were found during the measurement processing. These included an apparent shift between the model and measurement influence lines Figure 54, as well as differences in the structural behaviour Figure 55. The twin girder model was fitted to the measurements by optimization to determine if the behaviour observed in the measurements can be captured by the physical model. The vector of structural parameters is $\theta_{\mathrm{s}} = \{c_1, K_v, K^{\mathrm{l}}_{\mathrm{r}2}, K^{\mathrm{r}}_{\mathrm{r}2}, K^{\mathrm{l}}_{\mathrm{r}3}, K^{\mathrm{r}}_{\mathrm{r}3}, K^{\mathrm{l}}_{\mathrm{r}4}, K^{\mathrm{r}}_{\mathrm{r}4}, K^{\mathrm{l}}_{\mathrm{r}5}, K^{\mathrm{r}}_{\mathrm{r}5}\}$.

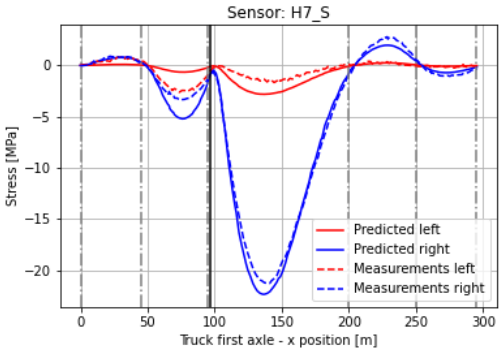


(a) Comparison for sensor H2.

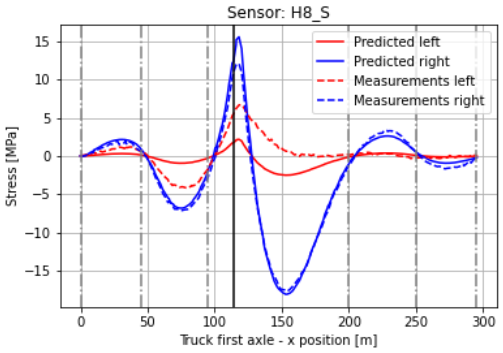


(b) Comparison for sensor H3.

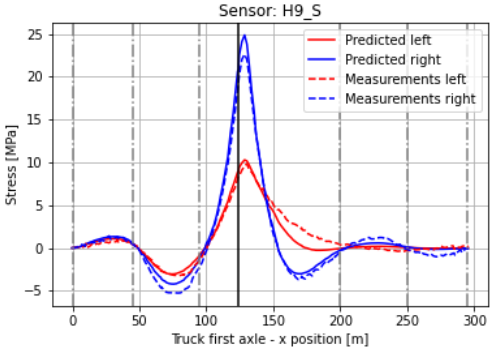
Figure 54: Shift in influence lines between model and measurement.



(a) Comparison for sensor H7.



(b) Comparison for sensor H8.



(c) Comparison for sensor H9.

Figure 55: Structural behaviour discrepancies between model and measurement.

Potential causes for these discrepancies include:

- Non-constant truck speed during the measurement runs.

- Incorrect assumptions in the measurement processing.
- Structural behaviour that can not be captured by the FE model.

C Appendix C - Implementation details

Python 3.8 and Julia 1.5 were used to perform the analyses presented in this report. The code is available upon request. A list of the main Python packages utilized is provided in the table below.

Table 24: Main Python packages used.

Package	Description	Use
NumPy	Fundamental package for computing in Python, providing basic functionality such as arrays, linear algebra, random number generation and efficient vectorized computation.	General computation.
SciPy	Scientific computing package for Python offering routines numerical integration, interpolation, optimization, linear algebra and statistics.	General computation.
Torch	Optimized tensor library for deep learning using GPUs and CPUs.	Likelihood evaluation.
Numba	Open source JIT compiler for Python and NumPy.	Transformation of Python code into fast machine code.
Numdifftools	Python package offering a suite of tools for numerical differentiation using an adaptive finite differences method.	Calculation of FE model gradients.
scikit-learn	Python library for supervised and unsupervised machine learning.	Implementation of kernel functions
PyJulia	Python interface to Julia.	Interface with the Julia FE model.
Dynesty	Pure Python implementation of Dynamic Nested Sampling.	Estimation of Bayesian posteriors and evidences.

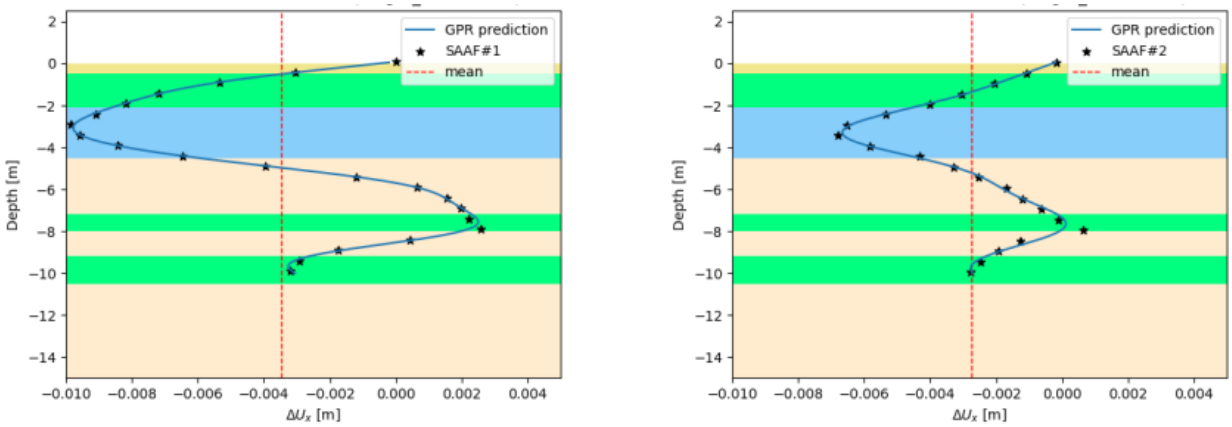
D Appendix D - Minimal examples

A number of minimal examples are presented in this section, with the aim of illustrating concepts that are relevant to Bayesian system identification when dependencies in the modeling uncertainties are considered, determining the impact of dependencies, and investigating the applicability of the approach detailed in Section 1.3. The following examples are presented:

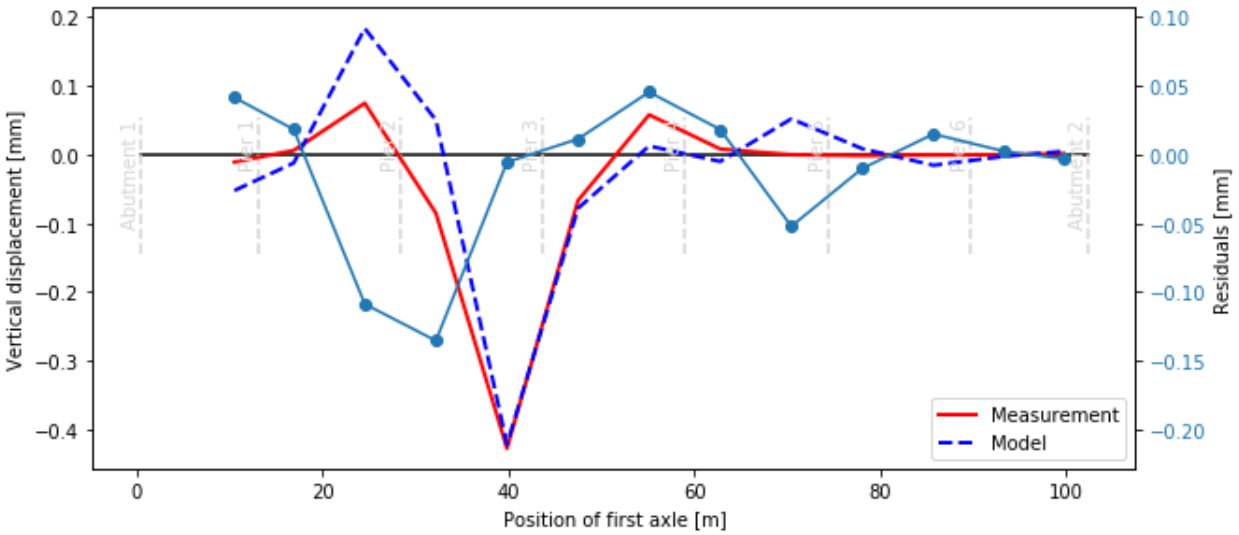
- Example D.1: Dependencies in model prediction uncertainties are illustrated in a synthetic sheet-pile wall example. Additionally, dependencies observed in the real-world case of bridge 705 are compared to those observed in a synthetic example.
- Example D.2: Inference of the correlation length for a simple synthetic case.
- Example D.3: Investigation of the effect of correlation on the posterior variance.
- Example D.4: Application of the FIM approximation method for selecting informative measurements for a simple beam.
- Example D.5: Comparison of the FIM approximation method with a method based on the KL-divergence.

D.1 Structure of dependence due to modeling uncertainties

As an initial step to determine the impact of dependencies in the model prediction error in Bayesian system identification, synthetic and real-world examples are studied in order to evaluate if such dependencies can be observed. Model predictions and measurements are presented for a synthetic sheet-pile wall example and a real-world concrete bridge in Figure 56. In Figure 56a, the predicted displacements along the height of a sheet pile wall are plotted against synthetic measurements. It can be observed that the discrepancies between model prediction have a specific structure and are not independently randomly distributed around the model prediction. This conclusion can also be drawn by Figure 56b. In this figure the real measured displacement influence line of bridge 705, a concrete bridge in Amsterdam where controlled loading tests were performed, is plotted along with the corresponding model prediction. The residuals appear to have a structure. These examples are indicative that modeling the residuals between model predictions and measurements as independent random variables may not be valid.



(a) Measurement and FE model prediction differences for a sheet pile profile pull-over test (from Chai (2019)).



(b) Measured influence line for a reinforced concrete slab bridge (bridge 705 in Amsterdam), model prediction and residuals (from Rozsas et al. (2020)).

Figure 56: Examples of dependencies in model prediction errors for synthetic sheet-pile wall (top) and real-world concrete bridge (bottom) cases.

D.2 Accounting for dependence in parameter estimation

In this example the feasibility of inferring the correlation length is explored. Realizations of five Gaussian random processes (GP's) with mean $\mu = 0$, marginal variance $\sigma = 1.0$ and exponential correlation (see Table 2) are generated over a domain with length $L = 295.0$. The domain is discretized into 165 points and a total of 50 realizations are generated for each $l_{\text{corr},0}$ in the set $\boldsymbol{l} = \{10.0,\ 20.0,\ 50.0,\ 100.0,\ 200.0\}$, with $l_{\text{corr},0}$ denoting the true value of the correlation length.

Bayesian inference is performed for each GP considering l_{corr} as an uncertain parameter with a uniform prior taken as $\mathcal{U}(10^{-3} \cdot l_{\text{corr},0},\ 5 \cdot l_{\text{corr},0})$. The posterior is calculated by numerical evaluation of Equation 1, using trapezoidal integration to obtain the evidence (Equation 16). The posterior mean per realization as a function of the number of observations from that realization used in the inference is shown in Figure 57. The variance of the posterior means is also plotted.

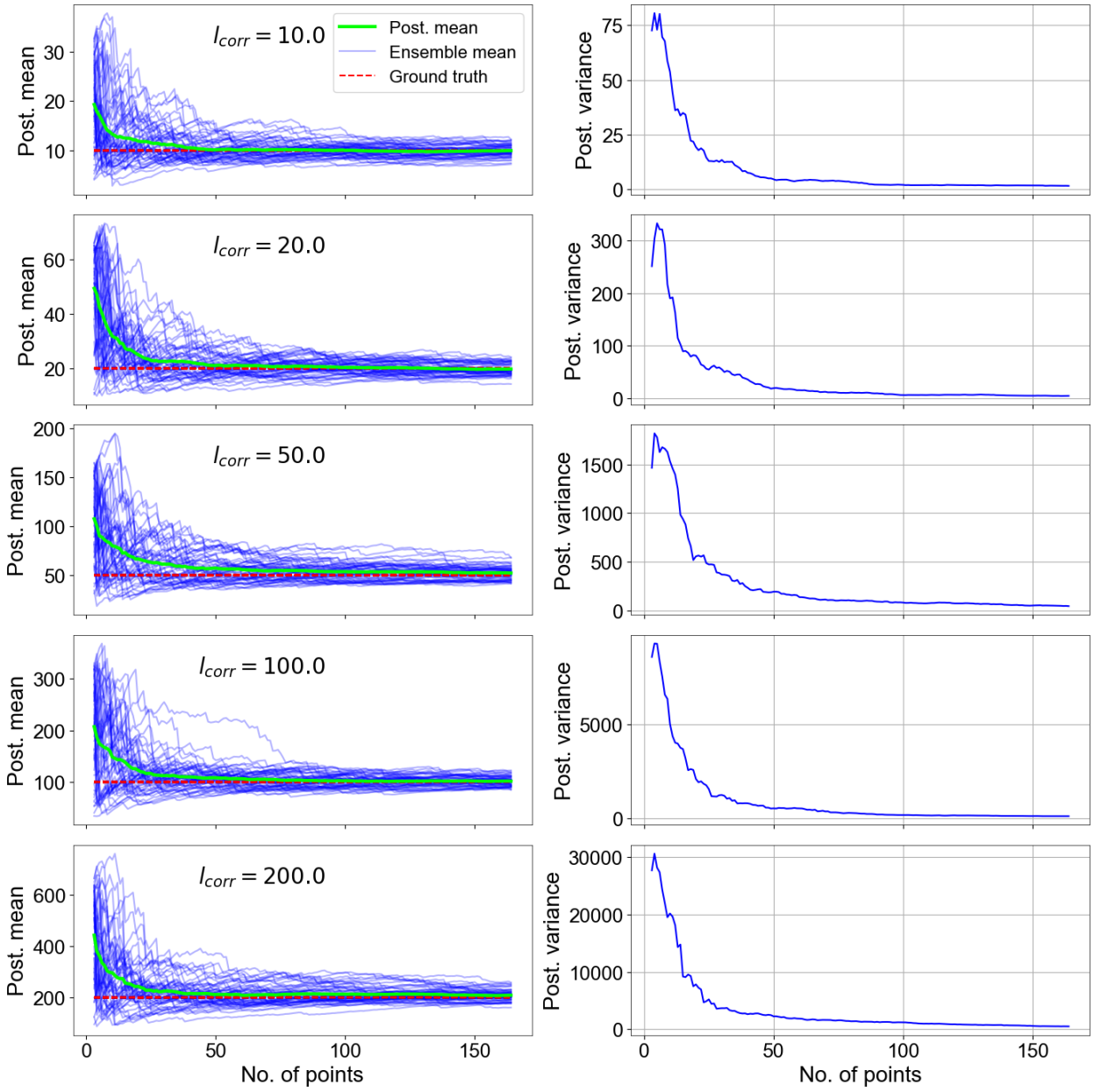


Figure 57: Posterior mean of l_{corr} per realization (left) and ensemble variance (right) as a function of the number of observations considered in the inference. The dashed red line denotes the ground truth, and the green line denotes the ensemble mean.

The size of the domain and the discretization are chosen to be similar to the IJsselbridge cases presented in Sections 6.2 and 6.3. Additionally, the true correlation lengths $l_{\text{corr},0}$ in the set $\boldsymbol{l}$ span the support of the uniform prior defined for the correlation length in those cases. This simple example serves to indicate that $N = 165$ points per realization is sufficient to estimate the mean of l_{corr} for $l_{\text{corr}} \in \boldsymbol{l}$ in the IJsselbridge case study.

D.3 The effect of dependence on posterior variance

The effect of dependence on the posterior variance is analyzed in this section using a simple example of a simply supported linear elastic beam, with the aim of illustrating how assumptions regarding the correlation can impact the results of Bayesian inference. The simply supported beam is loaded by a point load P at the midspan and is assumed to be outfitted with sensors measuring the vertical displacement y at $L/2$ and $L/4$, as shown in Figure 58.

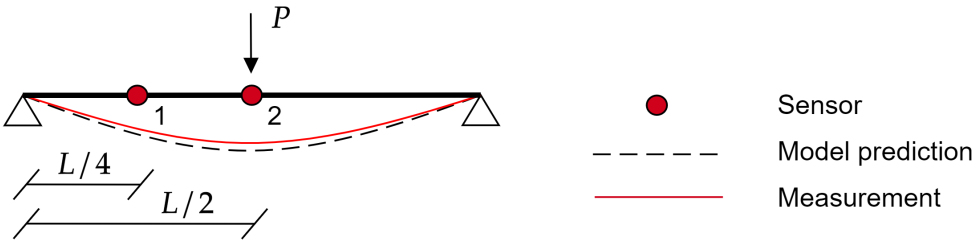


Figure 58: Illustration of simply supported beam with point load at the midspan and sensors at $L/2$ and $L/4$.

Bayesian inference is performed using the measured displacements to determine the posterior of the Young's modulus E. The prior distribution of E is defined as:

$$E \sim \mathcal{N}(\mu_0, \sigma_0) \tag{66}$$

The process generating the measurements is modeled as (capital letters denote random variables):

$$\boldsymbol{Y}_{\text{meas}} = \boldsymbol{y}_{\text{model}} + \boldsymbol{E}_{\text{model}} \tag{67}$$

where:

- $\boldsymbol{Y}_{\text{meas}}$ denotes the vector of measurements;
- $\boldsymbol{y}_{\text{model}}$ denotes the vector of the physical model prediction;
- $\boldsymbol{E}_{\text{model}}$ denotes the vector of model prediction error random variables.

The model prediction error $\boldsymbol{E}_{\text{model}}$ is taken as Normally distributed with zero mean and covariance:

$$\boldsymbol{\Sigma}(\rho) = \begin{bmatrix} \sigma^2_{\text{model}} & \rho \cdot \sigma^2_{\text{model}} \\ \rho \cdot \sigma^2_{\text{model}} & \sigma^2_{\text{model}} \end{bmatrix} \tag{68}$$

, where ρ denotes the correlation coefficient. It is noted that the quantities used in the example are not chosen to represent a realistic case. The properties and loading condition of the beam are as follows:

- Length $L = 20.0$ m;
- Load $P = 1000.0$ kN;
- Area moment of intertia $I_y = 5.538 \cdot 10^{-3}$ m^4;
- $E_{\text{true}} = 250.0$ GPa;
- $\mu_0 = 375.0$ GPa;
- $\sigma_0 = 125.0$ GPa;
- $\sigma_{\text{model}} = 2.5 \cdot 10^{-2}$ m.

Synthetic measurements using the true Young's modulus are generated. Bayesian inference is performed by numerical evaluation of Equation 1 for values of the correlation coefficient in the range $[0.0, 1.0]$ using a single measurement from each of the sensors. The prior and posterior distributions select values of ρ are shown in Figure 59.

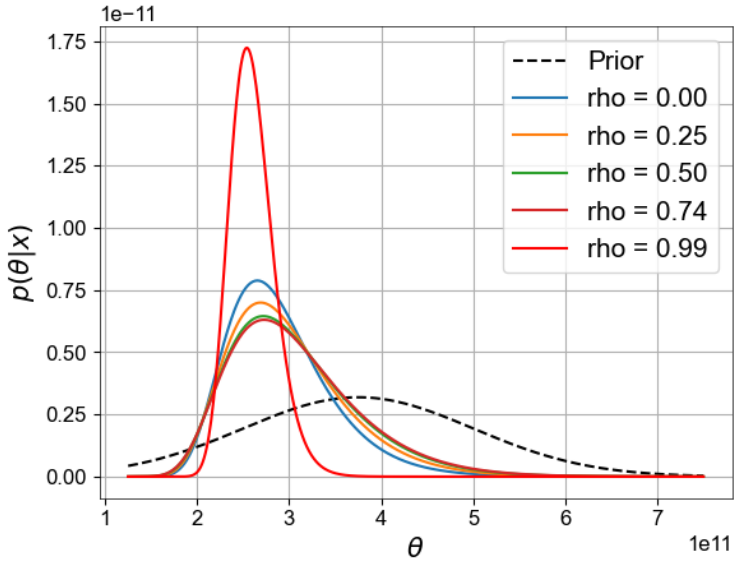


Figure 59: Posterior distribution calculated for varying correlation coefficients.

As expected, the posterior variance increases as the correlation coefficient increases. However, for $\rho = 0.99$ a significant reduction of the posterior uncertainty is observed. This counter-intuitive result is detailed in Simoen et al. (2013), and discussed in this section in order to

determine the impact of correlation in the model prediction on the posterior variance. The posterior variance as a function of ρ is plotted in Figure 60.

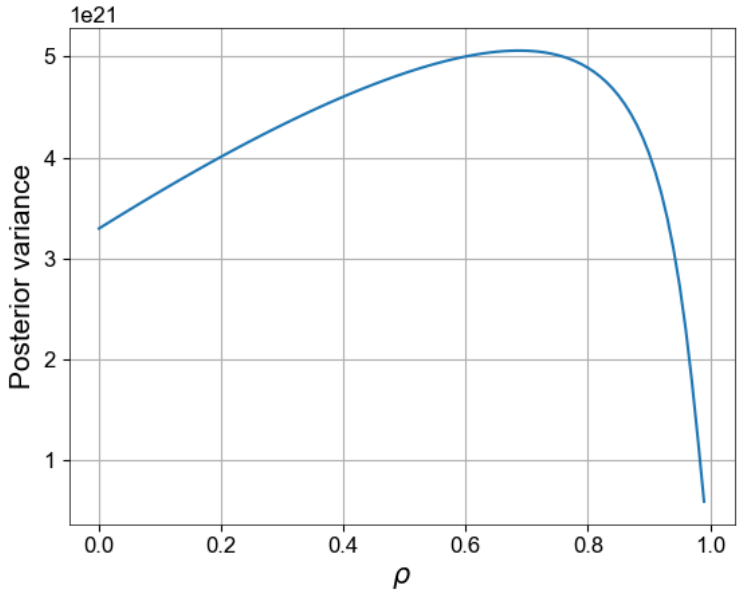


Figure 60: Posterior variance as a function of the correlation coefficient.

It is evident that the posterior variance is not a monotonic function of ρ, and higher values can lead to a sharper posterior distribution. A visual illustration of this effect is provided in Figure 61. The values of the deflections y_1 and y_2 for the prior support of E are calculated and the likelihood surface for each pair of possible observations is plotted for increasing ρ. The red line denotes the pairs of y_1 and y_2 predicted by the model for increasing values of E.

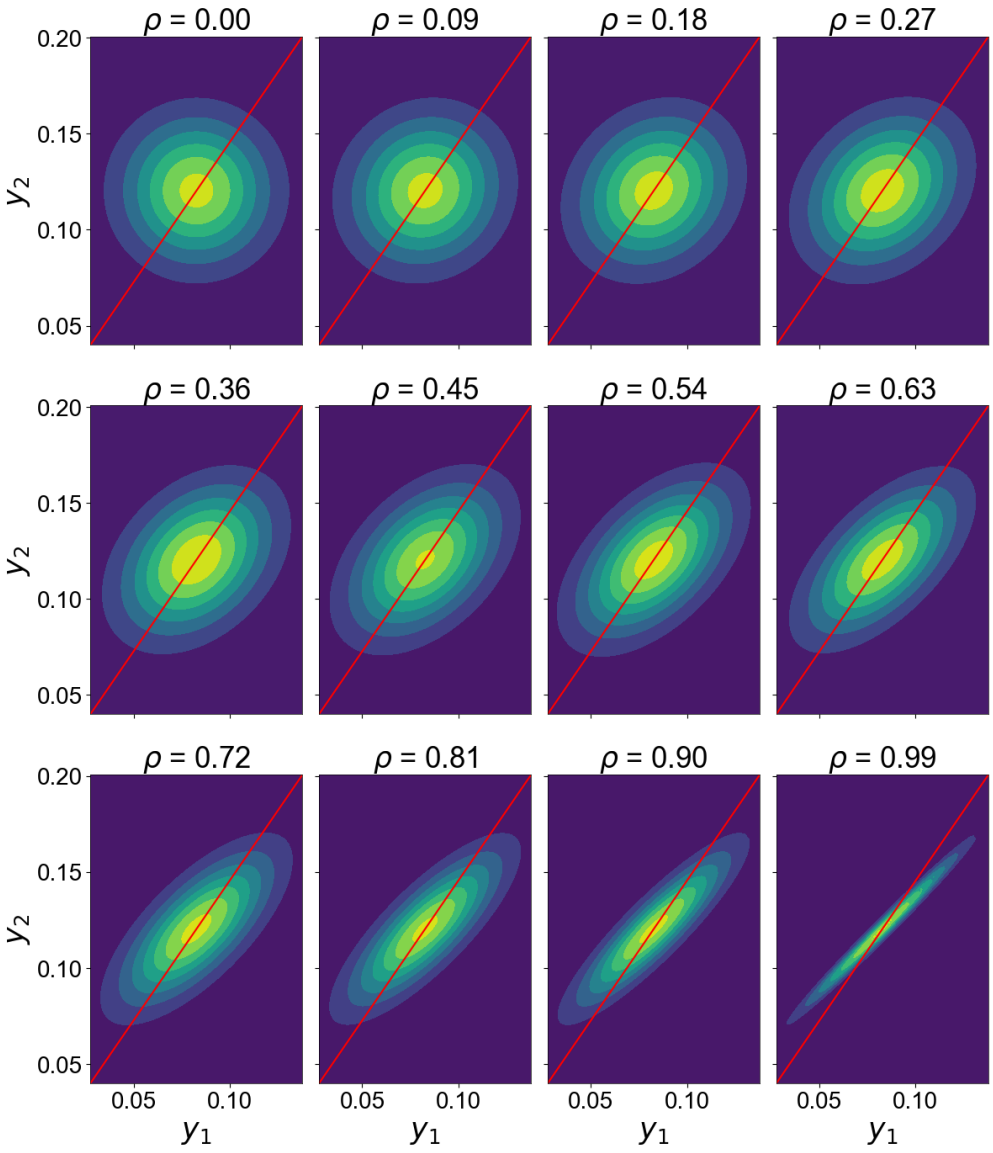


Figure 61: Visual explanation of the effect of increasing the correlation factor on the posterior variance.

The intersection of the vertical plane defined by this line with the likelihood surface will be the likelihood as a function of E, since there is a one-to-one correspondence between E and the model outputs: $(y_1, y_2) = f(E)$. It can be seen that for $\rho \geq 0.7$, the intersection between likelihood surface and the line defining the possible pairs of model outputs becomes narrower, with values outside the narrow intersection having low likelihood. Given that the posterior is dependent on the shape of the likelihood (assuming a wide prior distribution), this also results in the narrower posterior distributions observed previously. This example demonstrates that the relation between the correlation and the posterior variance is not monotonic, and that higher correlation does not necessarily results in wider posteriors and higher uncertainty. This highlights the importance of considering dependencies in the model prediction uncertainties in Bayesian system identification.

D.4 Calculation of the information entropy with the FIM approximation method.

A simply supported Euler-Bernoulli beam model is used as a toy example to study the FIM approximation method for measurement selection and to gradually introduce related concepts. For a point load P acting at point x_p of the beam, the displacement measured by a sensor at position x_d can be calculated as:

$$\frac{Pbx_d}{6EIL} \cdot (L^2 - b^2 - x_d^2), \tag{69}$$

where:

- P is the load
- b is the distance from the right support
- x_d is the sensor longitudinal position
- E is the Young's modulus
- L is the total length of the beam

The vertical deflection influence line for position x_d is obtained by calculating the deflection for a number of discrete locations of the point load across the structure. In practical applications of this method, a vector of nominal parameters θ_0 is defined based on previous knowledge and/or engineering judgement. In this case the Young's modulus E is considered as the unknown parameter and the the nominal value $E_0 = 2.1 \cdot 10^9$ Pa is assumed. The sensor position is taken as $x_\mathrm{d} = L/2$ The remaining parameter values are taken as shown below. It is noted that these values are arbitrary and were not chosen to be representative of a real system.

- $P = 10^{-4}$ N
- $I_y = 10^{-3}$ m^4
- $L = 10.0$ m

The correlated additive modeling uncertainty with i.i.d. Gaussian measurement uncertainty presented in Section 3.2.1 is used for the example. It is assumed that a radial basis function (Table 2) describes the correlations between different measurement positions. The uncertainty parameters are as follows:

- $\sigma_\mathrm{model} = 0.01$ m
- $\sigma_\mathrm{meas} = 0.01$ m
- $l_\mathrm{corr} = 1.0$ m

A significant advantage of the FIM approximation method over the KL-divergence method is the reduction of computational cost. It can be seen by the sketch of the algorithm in Section 3.5.2 that application of forward sequential sensor placement would require performing a full system identification for each load position at every iteration. This would be prohibitively expensive. Alternatively, global optimization could be applied, however, the computational cost would still be significant. In comparison, the approximate method only requires calculation of the gradient of the physical model with respect to the set of structural parameters $\boldsymbol{\theta}_\mathrm{s}$. These can be evaluated numerically at a fraction of the time of a full Bayesian inference. Furthermore this calculation is only performed once and no model evaluations are required between steps or between iterations. The gradient of the vertical beam displacement v w.r.t. the unknown parameter E is shown below:

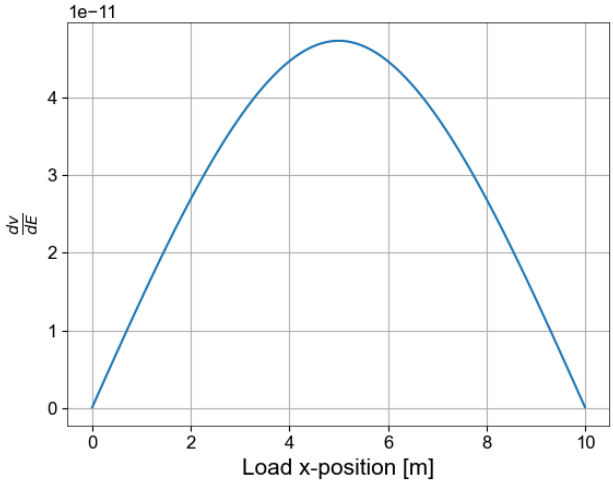


Figure 62: Gradient of the vertical deflection influence line w.r.t. E for sensor at position $x_\mathrm{d} = L/2$.

Initially no measurements are selected, i.e. $N_0 = 0$. For the first step Equation 28 is applied for every load position along the beam and the determinant of the FIM, $\det \boldsymbol{Q}$, is calculated. The optimal measurement is found at the position where $\det \boldsymbol{Q}$ is minimized, which corresponds to the highest information as shown by Equation 27. In subsequent steps the previously selected measurements are included in the observation matrix $\boldsymbol{L}$. This is repeated and an additional measurement is selected at each step until $N_i < N_\mathrm{max}$. The first four steps are illustrated in Figure 63. For each load position the calculated $\det \boldsymbol{Q}$ is plotted, forming a continuous curve along the length of the beam.

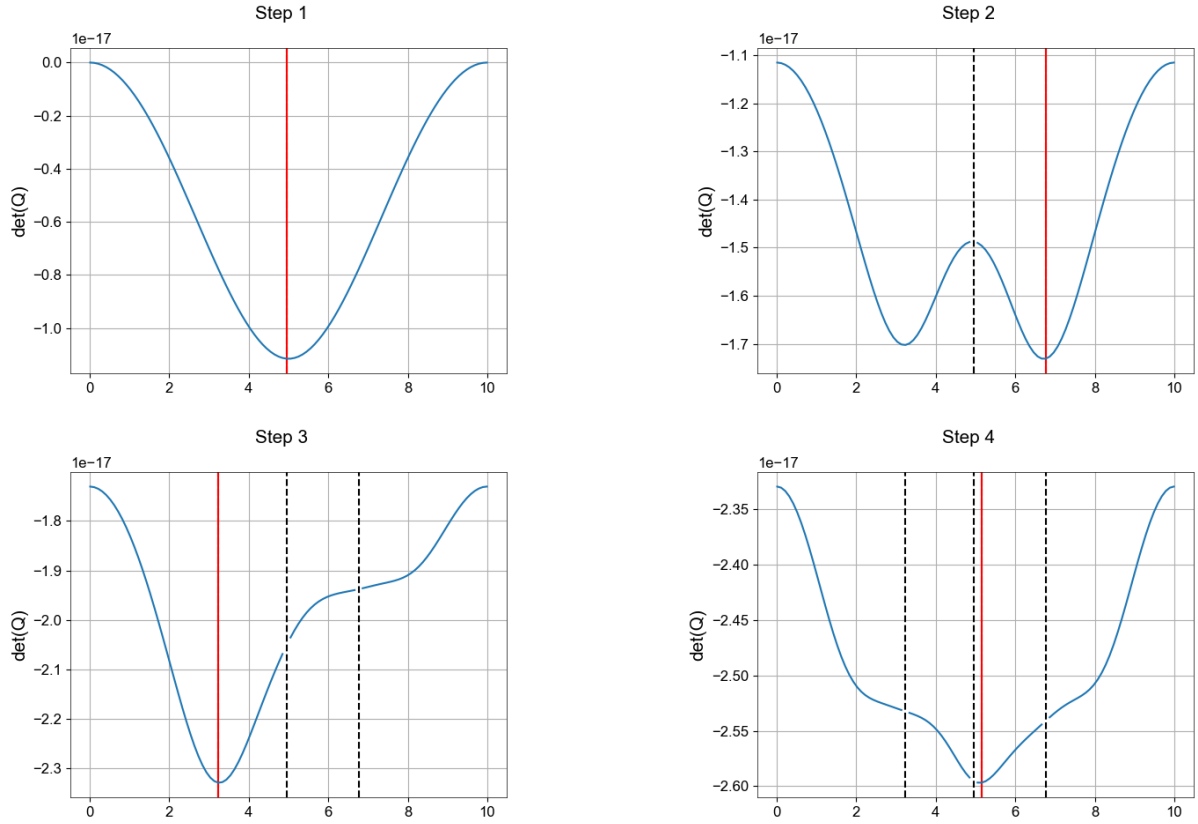


Figure 63: Successive steps of the measurement selection procedure using the FIM approximation method for a synthetic example.

As expected the measurement at the center of the beam is selected initially, with subsequent steps selecting measurements at a distance of approx. 1.0 m. This distance is determined by the covariance matrix Σ_0 included in Equation 28. It is evident by the shape of the $\det \boldsymbol{Q}$ curve that the correlation between closely spaced measurements results in a loss of information near the center. Since the gradient tends to zero near the boundaries and the information content of measurements diminishes near $x = 0$ and $x = L$, the selection favours points between the boundary and the already selected center point for the second and third step, in accordance with expectations. For the fourth step the optimal point is again near the center. The observed asymmetry of the $\det \boldsymbol{Q}$ curve is caused by the load positions being asymmetric with respect to the vertical axis passing through $L/2$.

D.5 Evaluation of the FIM approximation method for the IJsselbridge case

D.5.1 Overview

The aim of this example is to determine the applicability of the FIM approximation method for measurement selection in the case of the IJsselbridge. To this end the effect of the considered physical model parameters in the measurement selection is examined initially. This is done in order to confirm experimentally that the selected measurements are the most

informative for inferring the parameters of interest. Additionally, the impact of the assumed uncertainty parameters is also evaluated in a similar manner by performing measurement selection under the assumption of multiplicative modeling uncertainty and additive i.i.d. Gaussian noise. A Radial Basis correlation function is chosen to describe the unknown correlation structure and the covariance matrix is obtained for varying values of the correlation length parameter l_{corr}. It is shown that the spacing of the selected measurements corresponds to the assumed length scale. Finally, the approximate FIM method is compared with measurement selection using the exact solution of the KL-divergence, obtained from Bayesian system identification.

The FIM approximation method for measurement selection is applied to the case of the IJsselbridge. Additionally, the effect of the number of physical model parameters N_θ in relation to the number of measurements N_0 is discussed in Papadimitriou and Lombaert (2012). Specifically, it is noted that when $N_\theta > N_0$, the FIM $\boldsymbol{Q}$ becomes singular and $\det \boldsymbol{Q} = 0$. To remedy this, $\det \boldsymbol{Q}$ is instead calculated as the product of the non-zero eigenvalues of $\boldsymbol{Q}$. This allows for calculation of the optimal sensor positions with the FSSP algorithm starting with no selected measurements for any number of parameters.

D.5.2 Effect of physical model parameters

We consider the single girder FE model of the IJsselbridge presented in Section 5.3, parametrized by the LLF coefficient c_1 and the rotational stiffnesses of the first four supports K_{ri} with $i \in \{1, 2, 3, 4\}$. Measurement selection is performed using a global optimization algorithm[9]. The nominal values assumed for the physical parameters are:

- $c_1 = -0.175$
- $K_{\text{r1}-4} = 10^5$ kNm/rad

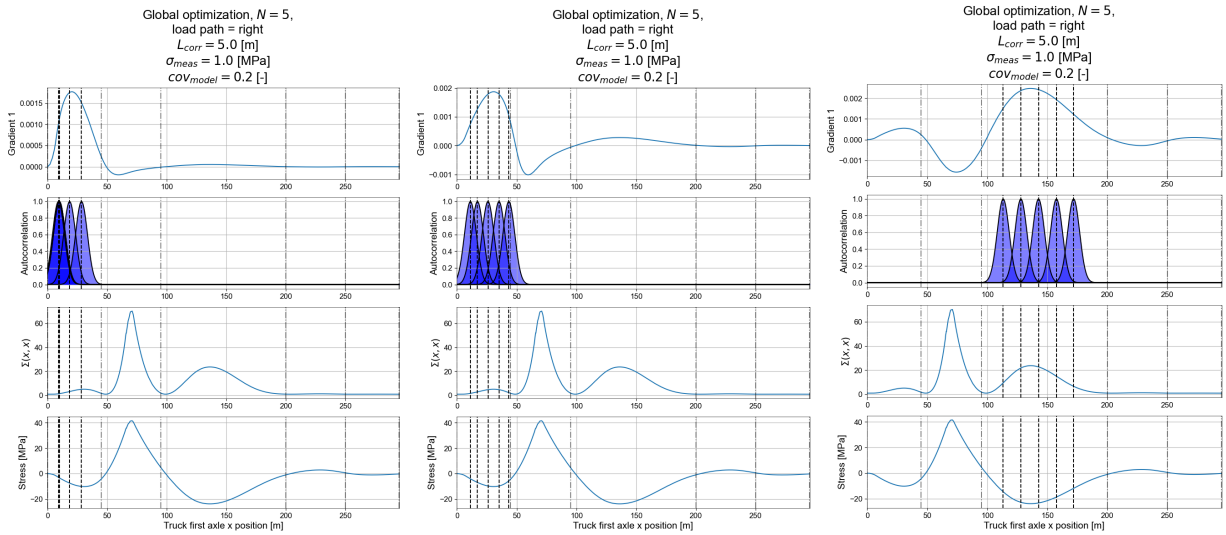
Three tests are performed for $N_{max} = 5$. For simplicity only the output for load on the right lane of the IJsselbridge is considered. For each test a different rotational spring stiffness K_r is considered in the parameter vector $\boldsymbol{\theta}$:

- Test 1: $\boldsymbol{\theta} = K_{\text{r1}}$
- Test 2: $\boldsymbol{\theta} = K_{\text{r2}}$
- Test 3: $\boldsymbol{\theta} = K_{\text{r3}}$

The results of the measurement selection are plotted in Figure 64. From left to right the three plots illustrate the results for $\boldsymbol{\theta} = K_{\text{r1}}$, $\boldsymbol{\theta} = K_{\text{r2}}$ and $\boldsymbol{\theta} = K_{\text{r3}}$. The gradient of the model output w.r.t. each parameter, the normalized autocorrelation and the diagonal of the covariance matrix, as well as the stress influence line are shown. The vertical dashed lines

[9]The results with FSSP and global optimization were compared. The observed differences were found to be negligible for the purpose of this example and the comparison is omitted.

represent the positions of the selected measurements while the grey dashed-dotted lines mark the locations of the supports. The assumed uncertainty parameters are also provided in the plot.



(a) Measurement selection for $\boldsymbol{\theta} = K_{\mathrm{r1}}$.

(b) Measurement selection for $\boldsymbol{\theta} = K_{\mathrm{r2}}$.

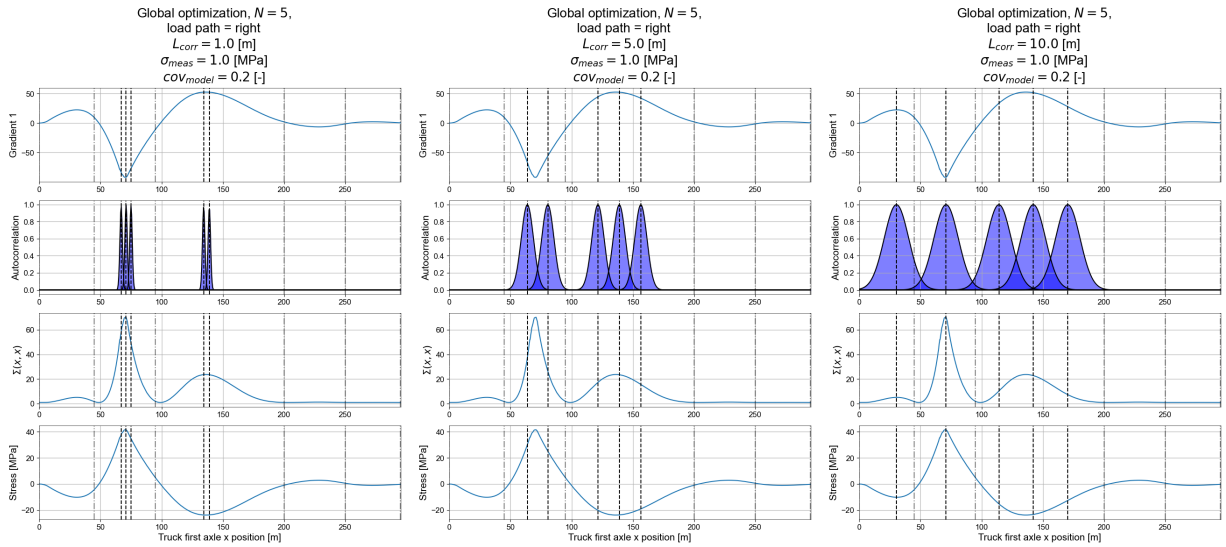
(c) Measurement selection for $\boldsymbol{\theta} = K_{\mathrm{r3}}$.

Figure 64: Illustration of the effect of considering different structural parameters in the gradient vector on the measurement selection.

It is immediately evident that the positions of the selected measurements depend on vector of parameters. In accordance with expectations, when the gradient vector includes the rotational stiffness of support F the selected measurements are located near that support. When the second and third rotational stiffness are considered the measurements are located near supports G and H respectively. These results indicate that we can expect the FIM approximation method to determine the locations of the optimal measurements that are most informative regarding the unknown parameters of interest.

D.5.3 Effect of uncertainty parameters

A similar experiment is performed to determine how the uncertainty parameters affect the optimal measurement positions. The vector of nominal parameters $\boldsymbol{\theta}_0$ is the same as that of the previous experiment. Three tests are performed for $N_{\max} = 5$ with varying values of the correlation length parameter $l_{\mathrm{corr}} = 1, 5$ and 10 m. The rotational stiffnesses are not included in the gradient vector ∇_θ and only the gradient of the stress influence line w.r.t. the LLF coefficient c_1 is calculated. The resulting measurement selection for increasing correlation length is shown in the figure below.



(a) Measurement selection for $l_{\text{corr}} = 1.0$ m.

(b) Measurement selection for $l_{\text{corr}} = 5.0$ m.

(c) Measurement selection for $l_{\text{corr}} = 10.0$ m.

Figure 65: Illustration of the effect of varying the correlation length parameter on the measurement selection.

The previous plots indicate that the correlation length has significant influence on the selected optimal measurement locations. Higher correlation lengths will result in reduced information content for closely spaced measurements. This leads to larger spacing of the optimal measurement locations. Furthermore, it can be expected that the other uncertainty parameters, modeling uncertainty coefficient of variation and measurement uncertainty standard deviation, as well as the kernel function used to describe the structure of the correlation, will all affect the optimal measurement locations.

D.5.4 Comparison with exact method

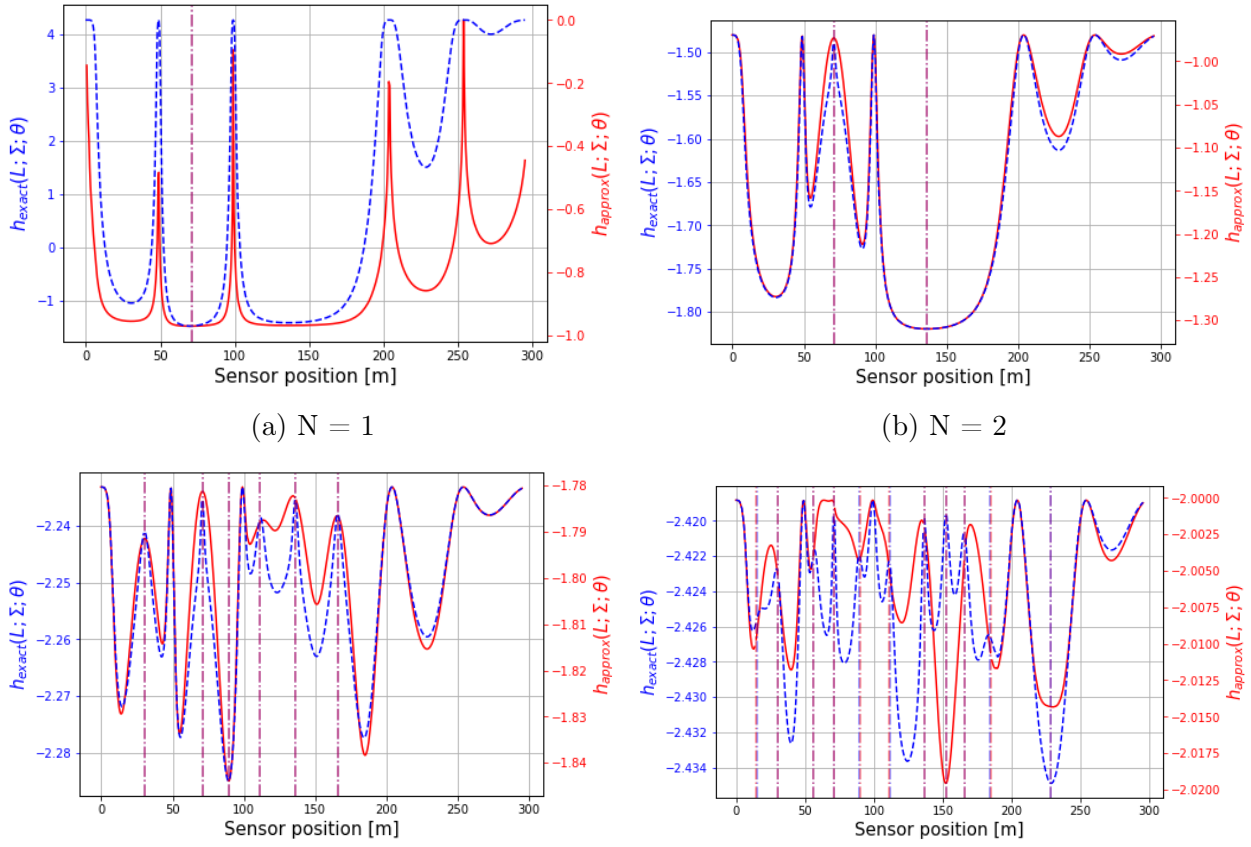
Results using the FIM approximation method are compared with the exact entropy calculation, referred to as the exact method. For the exact method, the full expression of the information entropy given in Equation 26 is used. At each step of every iteration, a full Bayesian inference is performed for the exact method using numerical integration. Inference by numerical integration is only feasible for a small number of parameters. Therefore only a single parameter, the lateral load coefficient c_1 is considered with a uniform prior distribution taken as $c_1 \sim U(-0.5, 0.5)$. In both cases the FSSP algorithm is applied to obtain the optimal measurement locations.

Synthetic measurements of the stress influence line are generated for a sensor located at $x = 65.37$ m, i.e. the location of sensor PRB-P4. A fine discretization with a maximum element length of 0.5 m is used in the IJsselbridge FE model for all calculations. The "ground truth" value for the LLF coefficient is $c_1 = -0.175$. It is assumed that the correlation in the

residuals can be described by a radial basis function. The uncertainty parameters are taken as:

- $\sigma_{\text{meas}} = 1.0$ MPa
- $C_{\text{v}} = 0.2$
- $l_{\text{corr}} = 10.0$ m

A comparison of the exact and approximate information entropy curves for each considered load position is shown in Figure 66. The information entropy curve is plotted for $N = 1, 2, 5$ and 10 measurements in order to evaluate the accuracy of the approximation for different N.



(a) N = 1

(b) N = 2

(c) N = 5

(d) N = 10

Figure 66: Comparison of the exact and approximate information entropy for increasing numbers of selected measurements.

For the assumed parameter the selected measurements for the two methods overlap up to $N = 10$. Deviations are observed for $N > 10$. It is noted that although the calculated entropy values differ between the two methods, the overall shape of the curve which determines the measurement selection is in good agreement. Discrepancies between the two methods are observed at $N = 1$ for regions of the influence line where the magnitude of the model output

is smaller in comparison to the measurement uncertainty, e.g. for sensor positions in the first, fourth and fifth spans corresponding roughly to the $x = [0, 45.0]$ and $x = [200.0, 295.0]$. For $N = 5$ and $N = 10$ discrepancies at the locations of selected measurements are also observed, indicating that the deviation of the two curves increase as the number of measurements increases. A comparison of the selected measurements for $N = 40$ is shown in the following figure.

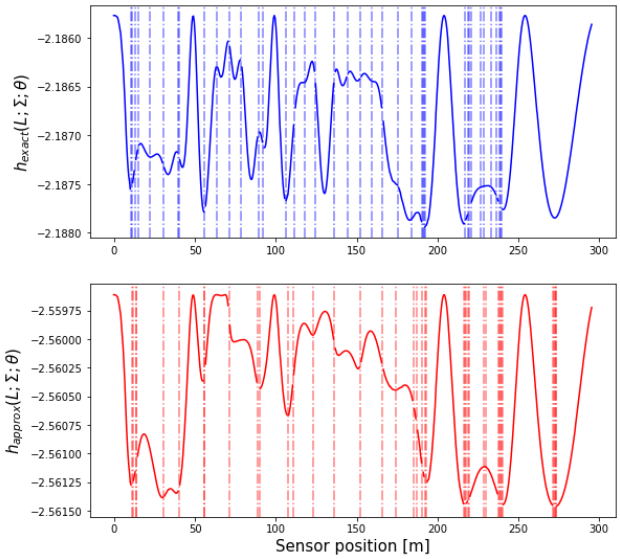


Figure 67: Comparison of the exact and approximate information entropy curves for $N = 40$.

It is evident that there are similarities in the overall layout for large numbers of measurements with clustering of measurements at roughly the same locations. Dissimilarities are also observed, notably the clustering of sensors in the peak located at roughly $x = 275.0$ m for the approximate entropy curve.

The experiments detailed in the previous sections suggest that the approximate FIM method is applicable for the IJsselbridge case. The parameters to be inferred as well as the structure of the correlations are taken into account in determining the optimal measurements and the measurement layout is sufficiently similar to the one resulting from the exact method despite discrepancies. It should be noted that these experiments are limited and aim to show the feasibility of this method for the case of the IJsselbridge. No general conclusions on the accuracy of the approximate method should be drawn from these results. A more systematic assessment of the approximate method should include:

- Investigation of the effect of measurement and modeling uncertainty.
- Investigation of the approximation accuracy when considering multiple structural parameters.
- Investigation of the effect of different correlation structures.
- Quantification of the accuracy of the approximation.

- Evaluation of the approximation when extended to sensor placement optimization.