



Probabilistic system identification and reliability updating for hydraulic structures – Application to sheet pile walls



Xuzheng Chai

August 2019



Graduation thesis to fulfill the requirement of
Master of Science in
Structural Engineering (Structural Mechanics) in
Faculty of Civil Engineering and Geosciences at
Delft University of Technology

Committee members:

Max Hendriks (Chair)	TU Delft
Timo Schweckendiek	TU Delft
Yuguang Yang	TU Delft
Árpád Rózsás (Daily supervisor)	TNO
Arthur Slobbe	TNO
Ana Teixeira	Deltares

Contents

Contents	2
Abstract.....	5
Acknowledgement	7
Glossary	8
List of acronyms and abbreviations	10
List of figures.....	11
List of tables.....	15
1 Introduction.....	16
1.1 Motivation	16
1.2 Research questions	17
1.3 Approach	19
1.4 Scope and limitations	21
1.5 Contributions.....	21
1.6 Thesis structure	22
2 Literature review	23
2.1 Overview	23
2.2 Structural health monitoring – Data collection	23
2.2.1 Definition	23
2.2.2 Examples of application.....	24
2.3 Structural health monitoring – Evaluation	24
2.3.1 System identification	24
2.3.2 Bayesian computational methods	26
2.3.3 Surrogate model.....	26
2.4 Assessment methods	27
3 Methods and tools	33
3.1 Overview	33

3.2	Surrogate models.....	33
3.2.1	Gaussian process.....	36
3.2.2	Model construction of Gaussian process regression.....	38
3.2.3	Prediction by Gaussian process regression.....	41
3.2.4	Adaptive fitting of Gaussian process regression model.....	42
3.2.5	Stopping criterion.....	44
3.2.6	Pseudo code.....	46
3.2.7	Illustrative example.....	48
3.2.8	Influence of noise.....	49
3.3	Uncertainty.....	51
3.3.1	Measurement uncertainty.....	51
3.3.2	Model uncertainty.....	53
3.3.3	Combined uncertainty.....	55
3.3.4	Correlation length.....	56
3.4	Bayesian inference.....	56
3.4.1	Parameter estimation.....	56
3.4.2	Overview of computation algorithms.....	58
3.4.3	Nested sampling.....	59
3.4.4	Information content.....	64
3.5	Reliability assessment.....	66
4	Case study: Sheet pile wall – modelling.....	68
4.1	Overview.....	68
4.2	Physical model.....	68
4.3	Synthetic observations.....	71
4.4	Bayesian inference of soil parameters.....	75
4.4.1	Soil properties to be inferred.....	75
4.4.2	Uncertainties.....	76

4.4.3	Correlation in physical model uncertainty	77
4.5	Construction of surrogate models	81
4.6	Overview of analyses	86
5	Case study: Sheet pile wall – results	87
5.1	Subcase 1: E_{top} & E_{mid}	87
5.1.1	Parameter identification	88
5.1.2	Reliability assessment	89
5.2	Subcase 2: E_{top} , ϕ_{top} & E_{mid} , ϕ_{mid}	90
5.2.1	Parameter identification	91
5.2.2	Optimal sensor selection	93
5.2.3	Influence of construction stages	95
5.2.4	Reliability assessment	97
5.2.5	Intermediate conclusions	98
5.3	Subcase 3: E_{top} , ϕ_{top} , Υ_{top} , C_{top} & E_{mid} , ϕ_{mid} , Υ_{mid} , C_{mid}	99
5.3.1	Optimal sensor selection	101
5.3.2	Reliability assessment	114
5.4	Discussion	125
6	Conclusions and recommendations	127
6.1	Answers to main research question	127
6.2	Answers to sub research questions	128
6.3	Recommendations	131
	References	132
Annex A	Used Python packages	139
Annex B	Sheet pile wall structures	140

Abstract

Given its geographical location and history, water defense is of utmost importance for the Netherlands. Structural Health Monitoring (SHM) offers a promising approach for system identification of hydraulic structures in this water defense system. The aim of SHM is to set sensors on structures and use the monitored responses to identify structural parameters of interest. However, many pertaining questions are unanswered concerning realistic hydraulic structures and monitoring systems: What type of sensors (e.g. strain gauge, SAAF, etc.) can and should be used in the monitoring system? How many sensors are needed, where and when to install these sensors? What is the influence of construction stages of structures on system identification? Considering the evaluation of a structure, what is the influence of system identification as well as construction stages on reliability (failure probability) of structures? Considering practical implementation: which computational algorithm is suitable and feasible? How to construct a proper surrogate model of the mechanical model to reduce computational time?

To answer these questions, a single anchored sheet pile wall is studied using a probabilistic approach. The sheet pile wall is modeled using the finite element (FE) method, synthetic data are used and Bayesian approach is adopted to cope with measurement uncertainty and model uncertainty. The information conveyed by sensors is quantified by the Kullback–Leibler (KL) divergence between prior and posterior distributions. Moreover, the correlation in model uncertainty of various structural responses is quantified by comparing a full-scale experiment from the literature and a corresponding calibrated 3D finite element model.

The results show that:

- A combination of different sensor types (in our case they are SAAF and strain gauge) should be used in the monitoring system (e.g. the combination of four different types of sensors outperforms the strain sensors on the sheet pile wall by conveying 40% more information with respect to the former);
- Even limited number of sensors can convey sufficient information. In our case, 3 sensors placed at proper locations can convey 90% information carried by 6, 8 and 9 sensors considering different responses. They should be installed as early as possible;

- The failure probability computed using posterior from system identification largely decreases compared with that computed using prior (the ratio of prior and posterior failure probabilities can go up to 10^{15} in our case);
- Delay of the start of monitoring during the construction stages decreases the information conveyed by sensors in system identification (the conveyed information can decrease by 50% in our case) and increases the computed failure probability in reliability analysis: the ratio of prior and posterior failure probabilities can be as large as 10^{30});
- MultiNest performs well in Bayesian inference in high dimensional problems;
- Gaussian process regression (GPR) with anisotropic radial basis function (RBF) kernel and white kernel as well as an adaptive infilling criterion is capable of constructing an accurate surrogate model even when it goes to high dimensionality. The error of surrogate model prediction can be explicitly explained.

To my knowledge the work presented in this thesis is the first application of combined system identification and reliability assessment for hydraulic structures, and the first detailed analysis of the effect of sensor installation time on system identification and structural reliability of hydraulic structures.

The findings imply that probabilistic system identification is a promising approach to substantially reduce our uncertainty in modelling hydraulic structures and in turns to increase their calculated safety. The approach has the potential to extend the working life of aging hydraulics structures and save costly strengthening and replacement. The analysis framework can also be applied to other structures in civil engineering.

Acknowledgement

I am very much grateful to my daily supervisor Árpád Rózsás in TNO, who kindly provided me with help for all the difficulty I met during the graduation project. He patiently and clearly taught me on probabilistic knowledge, programming, presentation, writing, etc. These are the most precious treasure I gained during my study in the past two years. It is my pleasure and my honor to work with him and to learn from him.

I would also like to thank Arthur Slobbe. He gave me a lot of advice on my work as well as my thesis writing at each progress meeting. I am thankful to Ana Teixeira for her advice related to geotechnical engineering and numerical setting of Plaxis. I would like to thank Max Hendriks, Timo Schweckendiek and Yuguang Yang to join my committee. They gave me valuable feedback on my results after every meeting and suggestions/comments on my thesis draft.

I am also grateful to work in the Department of Structural Reliability in TNO, people here are all friendly and helpful. I would like to address special thanks to Diego Allaix, the conversations with him are really helpful and appreciated. I would also like to thank the support of Deltares, which kindly provides the laptop, the Plaxis license and the access to their powerful computer. The developers of the open-source applications, e.g. Plaxis and Python and the researchers who made their work public are highly appreciated.

Many thanks to my friends with whom I enjoyed the two-year life in the Netherlands. The travelling with my friends over the Europe is a great memory and very much appreciated.

I would like to express the greatest thanks to my father, who brought me up and supports my study at TU Delft financially and mentally. With his encouragement I have no fear to move forward!

Glossary

The terminology and notation used here are based on the terminology accepted and used in (i) mathematical statistics and probability theory; (ii) structural reliability; (iii) geotechnical/hydraulic engineering. The roman numerals indicate precedence in case of conflicting terminology or notation in different fields. For clarity, the definition of some key terms are given here.

Coupled physical-probabilistic model: a model which contains both a physical model and a probabilistic model that are coupled, e.g. some or all parameters in the physical model are represented as random variables.

Distribution fitting: a part of statistical inference in mathematical statistics. In this study it is restricted to parameter estimation (inference) of probabilistic models.

Hydraulic structure: a load bearing structure that is fully or partially submerged in water and disrupts the flow of water.

Minimal problem: a problem that is as simple as possible yet able to capture the essential features of the examined question; “as simple as possible, but no simpler”. Minimal is used in the sense of minimal working example in programming.

Model: a mathematical representation of selected characteristics of an object or phenomenon.

Parameter estimation uncertainty: statistical uncertainty in the parameters of a given probabilistic model type, i.e. probabilistic model uncertainty is excluded.

Physical model (\mathcal{M}_{ph}): a deterministic model which describes/represents a physical phenomenon. Note that it can be empirical, first principles based, analytical (symbolic), numerical, etc. Provided with the same inputs it always yields the same outputs. Examples of physical models:

- a standardized, symbolic formula to calculate the shear resistance of a reinforced concrete beam;
- a nonlinear finite element model.

Probabilistic model (\mathcal{M}_{pr}): a model which describes/represents uncertainties. Here, we consider only approaches which comply with probability theory, e.g. fuzzy models are excluded.

Probabilistic model uncertainty: statistical uncertainty in the mathematical representation of the probabilistic model, i.e. model type.

Statistical inference: “the process of drawing conclusions about populations or other collections of objects about which we have only partial knowledge from samples” (*Simon, 1997*).

Statistical uncertainty: uncertainty stemming from finite sample size and in turns from sampling variability. We treat it as composed of two components: parameter estimation uncertainty and probabilistic model uncertainty

System identification: one of the key concepts in this study and referred to with many names in different disciplines and branches of engineering: in the geotechnical literature it is often termed *back analysis*, in aerospace and structural engineering *system* (or more particularly *damage*) *identification*. In structural reliability, although more encompassing, the term model (or Bayesian) *updating* is often used to refer to this problem. In almost all of the mentioned fields the term *inverse analysis* is frequently used. In this report the term *system identification* is decided to be used..

Synthetic data: data that is generated from a fully known model opposed to real data that is measured on a real system. The model used to generate synthetic data is often meant to represent a real system. In this work, synthetic data refers to data that is generated (simulated) from a finite element model and contaminated with random realizations from a known probabilistic model to add uncertainty.

List of acronyms and abbreviations

ACR	Area of credible region
COV	Coefficient of variation
DS	Directional simulation
FE	Finite element
FEM	Finite element method/model
FORM	First order reliability method
GA	Genetic algorithm
GP	Gaussian process
GPR	Gaussian process regression
HMC	Hamiltonian Monte Carlo
IS	Importance sampling
KL	Kullback–Leibler
KLD	Kullback–Leibler divergence
LH	Latin Hypercube
MCMC	Markov Chain Monte Carlo
MCS	Monte Carlo simulation
MH	Metropolis-Hastings
MSE	Mean squared error
NAP	Normal Amsterdam Level
RBF	Radial basis function
RE	Relative error
SHM	Structural health monitoring
SLS	Serviceability limit state
SORM	Second order reliability method
TMCMC	Transitional Markov Chain Monte Carlo
ULS	Ultimate limit state

List of figures

Figure 1.1: A schematic overview of the approach used in this study.	20
Figure 3.1: Kernels in Gaussian process regression [Duvenaud (2014)].....	37
Figure 3.2: Variance of RE with number of sampling points.	45
Figure 3.3: Variance of threshold with number of sampling points.	45
Figure 3.4: Initial sampling points and fit.....	48
Figure 3.5: Demonstration of predictive error (1).	49
Figure 3.6: Demonstration of predictive error (2).	49
Figure 3.7: Demonstration of predictive error (3).	49
Figure 3.8: Response without noise component.	50
Figure 3.9: Response with noise component.	50
Figure 3.10: RE without noise component.	50
Figure 3.11: RE with noise component.....	50
Figure 3.12: Showcase of nested sampling.....	60
Figure 3.13: Example of nested sampling.....	62
Figure 3.14: Showcase of degenerate distribution.....	63
Figure 4.1: Scheme of the case study: dimensions.	68
Figure 4.2: Cross-sectional properties of sheet pile [Teixeira et al. (2015)].	70
Figure 4.3: FE mesh a representative amplified, deformed shape of the analyzed sheet pile wall.	71
Figure 4.4: Strain of sheet pile wall corresponding to “real” values of soil properties.	72
Figure 4.5: Horizontal displacement of sheet pile wall corresponding to “real” values of soil properties.....	73
Figure 4.6: Vertical displacement of top ground corresponding to “real” values of soil properties.....	73
Figure 4.7: Top view of sheet pile to indicate strain gauge.	74
Figure 4.8: Layout of full sensor configuration.	74
Figure 4.9: Showcase of the selection of support.	76
Figure 4.10: A 3D view of the FE model of the pull-over test.	78
Figure 4.11: Pull-over force over time during the experiment.	78
Figure 4.12: Showcase of installed SAAF's.	78
Figure 4.13: SAAF#1 measured and Plaxis predicted values at 12:30:00.....	79
Figure 4.14: SAAF#2 measured and Plaxis predicted values at 12:30:00.....	79

Figure 4.15: SAAF#1 measured and Plaxis predicted values at 12:50:00.....	79
Figure 4.16: SAAF#2 measured and Plaxis predicted values at 12:50:00.....	79
Figure 4.17: Fitting the difference between SAAF#1 measured and Plaxis predicted values at 12:30:00.	79
Figure 4.18: Fitting the difference between SAAF#2 measured and Plaxis predicted values at 12:30:00.	79
Figure 4.19: Fitting the difference between SAAF#1 measured and Plaxis predicted values at 12:50:00.	80
Figure 4.20: Fitting the difference between SAAF#2 measured and Plaxis predicted values at 12:50:00.	80
Figure 4.21: Maximum of normalized prediction standard deviation of surrogate models.....	82
Figure 4.22: R^2 value of surrogate models.....	83
Figure 4.23: 1D visualization of S#20 at P7 _{start}	84
Figure 4.24: 1D visualization of G#35 at P4 _{start}	84
Figure 4.25: ε_{spw} at S#20 installed at P4 _{start}	85
Figure 4.26: ε_{spw} at S#45 installed at P4 _{start}	85
Figure 5.1: Sensor placement measuring the strain of sheet pile wall.	87
Figure 5.2: Prior distribution of Subcase 1.	88
Figure 5.3: Posterior distribution of Subcase 1.....	88
Figure 5.4: Failure region in prior distribution of Subcase 1.....	90
Figure 5.5: Failure region in posterior distribution of Subcase 1.	90
Figure 5.6: Prior distribution of Case 2.	91
Figure 5.7: Posterior distribution of Case 2.	91
Figure 5.8 Responses of sensors installed at P4 _{start} in Subcase 2.....	93
Figure 5.9: KLD matrix plot.	94
Figure 5.10: Variation of KLD with different number of sensors.	94
Figure 5.11: Posterior in Subcase 2 with 3 optimal sensors.	95
Figure 5.12: Posterior in Subcase 2 with 6 sensors.	95
Figure 5.13: KLD matrix plot with sensors installed at P4 _{start} ,	96
Figure 5.14: KLD matrix plot with sensors installed at P5 _{start} ,	96
Figure 5.15: KLD matrix plot with sensors installed at P6 _{start} ,	96
Figure 5.16: KLD matrix plot with sensors installed at P7 _{start} ,	96
Figure 5.17: Variation of KLD with sensor installation phase. The three optimal sensors are indicated in brackets.	97

Figure 5.18: Comparison between posterior with sensors installed at $P4_{start}$ and that with sensors installed at $P7_{start}$. The contour plots in the off-diagonal plots are 90% credible regions.....	97
Figure 5.19: Failure projections in prior distribution.....	98
Figure 5.20: Failure projections in posterior distribution.	98
Figure 5.21: Prior in Subcase 3.....	101
Figure 5.22: Sensor placement in Subcase 3-1.	102
Figure 5.23: KLD matrix plot with sensors installed at $P4_{start}$	103
Figure 5.24: KLD matrix plot with sensors installed at $P5_{start}$	103
Figure 5.25: KLD matrix plot with sensors installed at $P6_{start}$	103
Figure 5.26: KLD matrix plot with sensors installed at $P7_{start}$	103
Figure 5.27: Variation of KLD in Case 3-1.	104
Figure 5.28: The most informative posterior in Case 3-1.....	104
Figure 5.29: Posterior based on 6 strain sensors on sheet pile wall installed at $P4_{start}$	105
Figure 5.30: Sensor placement in Case 3-2.	106
Figure 5.31 KLD matrix plot with sensors installed at $P4_{start}$	107
Figure 5.32 KLD matrix plot with sensors installed at $P5_{start}$	107
Figure 5.33 KLD matrix plot with sensors installed at $P6_{start}$	107
Figure 5.34: KLD matrix plot with sensors installed at $P7_{start}$	107
Figure 5.35: Variation of KLD in Case 3-2.	108
Figure 5.36: The most informative posterior in Case 3-2.....	108
Figure 5.37: Sensor placement in Case 3-3.	109
Figure 5.38: KLD matrix plot with sensors installed at $P4_{start}$	110
Figure 5.39: KLD matrix plot with sensors installed at $P5_{start}$	110
Figure 5.40: KLD matrix plot with sensors installed at $P6_{start}$	110
Figure 5.41: KLD matrix plot with sensors installed at $P7_{start}$	110
Figure 5.42: Variation of KLD in Case 3-3.	111
Figure 5.43: The most informative posterior in Case 3-3.....	111
Figure 5.44: Optimal sensor placement at $P4_{start}$ in Case 3-4.	112
Figure 5.45: Optimal sensor placement at $P5_{start}$ in Case 3-4.	112
Figure 5.46: Optimal sensor placement at $P6_{start}$ in Case 3-4.	112
Figure 5.47: Optimal sensor placement at $P7_{start}$ in Case 3-4.	112
Figure 5.48: Variation of KLD in Case 3-4.	113
Figure 5.49: The most informative posterior in Case 3-4.....	113
Figure 5.50: Response with sensors installed at $P4_{start}$ in Case 3-4.	114

Figure 5.51: Prior and posterior failure probabilities in SLS for different sensor installation times.....	116
Figure 5.52: Projection of failure regions in prior.	118
Figure 5.53: Projection of failure regions in posterior with sensors installed at $P4_{start}$	119
Figure 5.54: Projection of failure regions in posterior with sensors installed at $P5_{start}$	119
Figure 5.55: Projection of failure regions in posterior with sensors installed at $P6_{start}$	119
Figure 5.56: Projection of failure regions in posterior with sensors installed at $P7_{start}$	119
Figure 5.57: Showcase of failure probability in sheet pile wall failure.	121
Figure 5.58: Showcase of failure probability in anchor failure.	123
Figure 5.59: Showcase of failure probability in structural failure.	124
Figure 0.1: Cross-sections through some typical permanent sheet-pile structures (Clayton, Woods, Bond, & Milititsky, 2014).	140
Figure 0.2: Anchored sheet pile wall schemes (Clayton et al., 2014).	141

List of tables

Table 2.1: Overview of previous works in system identification of hydraulic and geotechnical structures (the variables and acronyms are explained in Table 2.2). Studies with probabilistic system identifications where the uncertainty in the estimated parameters is quantified are highlighted with italic font.....	29
Table 2.2: Explanation of symbols used in Table 2.1.....	32
Table 3.1: Overview of surrogating tools.	35
Table 3.2: Overview of selected number of initial points.....	39
Table 3.3: A cursory overview of studies using local surrogating.	43
Table 3.4: Comparison between global surrogate and local surrogate.	44
Table 3.5: Pseudo code of the proposed adaptive Gaussian process surrogate method.	47
Table 3.6: Explanation of notation.....	51
Table 3.7: Comparison of different computational algorithms.....	59
Table 4.1: Soil properties.....	69
Table 4.2: Components in the FE model.	70
Table 4.3: Prior information.	75
Table 4.4: Standard deviation of each uncertainty components.	76
Table 4.5: Hyperparameters of the fitted GPR model.	80
Table 4.6: Overview of analyses.....	86
Table 5.1: Comparison of $E_{top} - E_{mid}$ between prior and posterior.....	88
Table 5.2: Prior and posterior failure probabilities in Subcase 1.....	90
Table 5.3: Comparison of $E_{top} - E_{mid}$ between prior and posterior.....	92
Table 5.4: KLD in Case 3-1.....	102
Table 5.5: Comparison of $E_{top} - E_{mid}$ between prior and posterior.....	105
Table 5.6: KLD in Case 3-2.....	106
Table 5.7: KLD in Case 3-3.....	109
Table 5.8: KLD in Case 3-4.....	112
Table 5.9: Failure probability in SLS.	115
Table 5.10: Failure probability in sheet pile wall failure.....	120
Table 5.11: Failure probability in anchor failure.	122
Table 5.12: Failure probability in structural failure.....	124
Table 0.1: Main features of different sheet pile wall materials.	142

1 Introduction

1.1 Motivation

Given its geographical location and history, water defense is of utmost importance for the Netherlands, e.g. 21% of its population lives under the sea level, about 60% of the Netherlands is flood prone, and about 55% of its GDP is in the 50 km zone of its coastline (*Slomp, 2012*). To further illustrate the importance of the water defense system: it is estimated that the flood of 1953 incurred about 5.4 billion euros damage (present day cost) and claimed 1836 lives (*Rijkswaterstaat*). On a European level: “The value of the economic assets within 500 m of the coastline is estimated at € 500-1000 billion. In addition, 35% (€ 3.5 trillion) of the total GDP of the 22 European coastal member states is generated in the area within 50 km of the coast, an area which hosts moreover 1/3rd of the EU population.” (*Union, 2009*).

Hydraulic structures, such as sheet pile walls, play an essential role in this water defense system, which are usually constructed to retain earth, water, or any other filling materials. Sheet pile walls (see Annex B) are usually steel-made and relatively thin, e.g. 12 mm, and hence susceptible to corrosion which can be undesirable aesthetically or compromise structural safety. In addition, most of these structures were built decades ago and are now approaching the end of their lifetime. Therefore, it is necessary to regularly inspect these retaining walls to assess if they still comply with safety requirements and to make decisions if they should be strengthened or even replaced.

The current approach to make such decisions is based on engineers’ inspection and judgement, which is very costly, labor-intensive and time-consuming considering the enormous size of hydraulic structures. For example, it could take hours or even days to inspect a several kilometers long dike. These inspections occur on a regular basis or when signs of problems are observed/reported (e.g. when problems experienced with cranes standing on the high side of sheet pile wall due to increased settlement). Neither the regular or the warning-triggered inspections lead to an optimal intervention time. Moreover, such inspection is highly dependent on the skills and experience of the inspector, and it is often limited by the accessibility of elements, e.g. foundations, which means some damage may remain hidden for a long time. Besides, the result of these inspections is not guaranteed to be accurate or perfect as humans could make mistakes, especially during time-consuming and repetitive tasks, which may lead to waste of labor and budget (e.g. some sheet pile walls still in good condition are replaced).

All these considerations constitute the motivation of automated structural health monitoring (SHM).

SHM offers an alternative and complementary approach to the current inspection techniques with reduced human labor. In a typical and complete SHM system, sensors are installed on a structure to collect data automatically, and then information is extracted that describes the response of the structure (e.g. deformation) and structural parameters (e.g. bending stiffness of a sheet pile) or soil parameters (e.g. elastic modulus). Another advantage of SHM is that since it is a real-time monitoring process, a warning system could be built on top of it based on collected data so that the abnormal behavior of structures, e.g. very large deformations, could be noticed quasi instantly. Moreover, considering the design and evaluation aspect, lower partial factor might be used due to reduced uncertainty gained by the SHM system. This has the potential to make structural design less conservative and thus may lead to a more economic design and maintenance.

1.2 Research questions

At present, there are hundreds of hydraulic structures in the Netherlands, and around 250 of them will have to be replaced in the coming decades and another 800 of them may no longer function properly in the future (*Lassing & Barneveld*). They are becoming vulnerable due to aging, erosion, etc. Thus, SHM could be used as an efficient tool to detect possible current damages, and in the long run, it could act as a warning system as mentioned above to dynamically provide warning on the onset of failure. To achieve this goal, system identification¹ based on data from monitoring system could be performed to locate the damages (e.g. (*Peeters, 2000*)), and reliability based methods are widely used to assess structures by computing the failure probability (e.g. (*Teixeira et al., 2015*)). However, there is seldom a combination of SHM-based system identification and reliability analysis. This knowledge gap is the motivation of this study and raises the main research question:

What is the effect of monitoring hydraulic structures on structural reliability? What is the influence of starting the monitoring at different construction stages?

Although SHM has been developed for decades, it is still not widely applied in practice and there are important challenges both in theory and practice. One common and ubiquitous

¹ It represents soil parameter identification in this thesis. Detailed description can be found in Glossary.

challenge in designing monitoring systems is the selection of sensor type and their placement (*Papadimitriou, 2004*). An optimized sensor configuration can largely reduce the number of needed sensors and thus save budget and labor. Although many methods exist to place sensors optimally in a SHM system, e.g. (*Papadimitriou & Lombaert, 2012; Yi, Li, & Gu, 2011*), and application of that in civil engineering, for instance a bridge example is presented in (*Meo & Zumpano, 2005*), there is limited information for hydraulic structures. Therefore, the following sub questions are considered for the SHM of hydraulic structures:

1. What type of sensors can be and should be used?
2. How many sensors are needed, where to put these sensors and when to install them (during construction or afterwards)?

After the data is collected by the sensors of a monitoring system, the assessment of the state of a structure (e.g. the determination of some structural parameters) is the most crucial aspect. This procedure is referred to as system identification, in which a physical model (e.g. finite element (FE) model) is included to describe the complex mechanical behavior of a typical civil engineering structure. The essence of system identification is to use the data from sensors to calibrate the physical model to minimize the difference between the measured values and model predictions (see e.g. (*Capecchi & Vestroni, 1999*)). However, by reviewing the system identification literature for hydraulic structures with focusing on requirements for practical applicability, the following limitations and/or challenges are identified (the details of the literature review with comprehensive references to the reviewed papers are presented in Chapter 2):

- Uncertainties (i.e. the residual difference after calibration) are often not considered (deterministic analysis rather than probabilistic) (see e.g. (*J. Deng, Lee, & sciences, 2001; Levasseur, Malecot, Boulon, & Flavigny, 2008*)), which are however inevitably present;
- When uncertainty is considered, observations are almost always considered to be independent, i.e. the effect of correlation is not explored (see e.g. (*J Zhang, Tang, Zhang, & Engineering, 2009; L. L. Zhang, Zhang, Zhang, & Tang, 2010*));
- In case of probabilistic analysis (Bayesian inference) typically only two parameters are identified, and the highest number is four, see e.g. (*Kelly & Huang, 2015; L. Wang,*

Hwang, Luo, Juang, Xiao, et al., 2013), more parameters are expected to be influential in practical applications;

Therefore, these limitations lead to the following sub question:

3. Which Bayesian computational method fits best in the context of probabilistic system identification of realistic hydraulic structures that are characterized by computationally expensive likelihood functions considering correlation, high dimensionality (i.e. more parameters are considered simultaneously) and multimodal and/or degenerate posteriors?

Also, to evaluate the performance of such inference, or equivalently to estimate the information conveyed by sensors, it is favorable to introduce a quantitative criterion as it makes evaluation more straightforward. And thus it results in this sub question:

4. How to quantitatively evaluate the information conveyed by sensors?

In practical implementation, the physical model is usually required to run thousands or tens of thousands times to perform Bayesian inference. This is too computational expensive and practically intractable, for example it would not make sense to make decisions after days of calculation if structural failure really happens. Therefore, a surrogate model, which is considerably cheaper in computation, could be constructed to replace the physical model and to support the extensive calculations. Since the surrogate modeling approaches in the hydraulic structures system identification literature cannot be scaled to higher dimensional cases – as they deal with low dimensional, simplified cases, we have to seek an answer to another sub question:

5. How to construct an accurate surrogate model of the physical model in a reasonable computation time?

1.3 Approach

In this study a probabilistic approach is selected to identify the non-directly observable soil parameters. In contrast with many hydraulic and geotechnical engineering studies we formulate the engineering problem in a rigorous mathematical form and adopt the language and methods widely used in mathematical statistics, probability theory, and information theory. The Bayesian paradigm is selected for representing, interpreting, and calculating probabilities. One of the main reasons for this is its compatibility with engineering and structural reliability way

of thinking about the world and models. A (physical) model based approach is selected that has at its core a numerical model that is based on our understanding of structural and soil mechanics. The other approach in SHM is a data-driven modelling (*Worden & Manson, 2007*), see for example to following applications (*Diez et al., 2016; Lang et al., 2008*). However, the data-driven approach is deemed inadequate in our context where we aim to gain insight into the structural behavior on the abstraction and modeling level of current engineering, e.g. to learn about soil properties which are not directly observed. A high level, schematic overview of the adopted approach is shown in Figure 1.1.

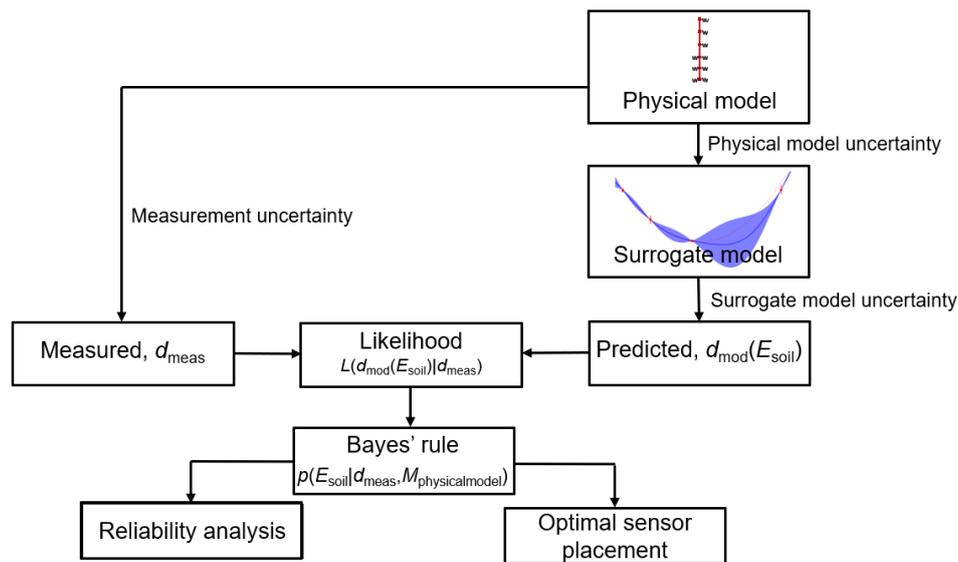


Figure 1.1: A schematic overview of the approach used in this study.

The core idea of this probabilistic approach is to treat the unknown parameters as random variables and make inference of them by considering measurement and model uncertainties, and use the updated information of the unknowns in reliability analysis. When observations are collected from sensors on real structures, measurement uncertainty appears due to the potential error of measuring devices, improper installation, and/or operation of such devices. Moreover, there is also model uncertainty that is the difference between FE models (model predictions) and real structures. If these uncertainties are taken into account, a joint distribution of the estimated parameters can be obtained through Bayesian inference. Then reliability analysis can proceed straightforwardly² using the joint distribution of the estimated parameters as inputs.

Therefore, three main steps are taken to develop this approach:

² Conceptually straightforward but numerically challenging as it requires the accurate estimation of the tail of the joint density function since that is expected to govern reliability.

- The construction of surrogate models of the monitored structural responses (i.e. replacement of FE model to reduce computational time).
- The estimation of the non-directly observed parameters, e.g. soil parameters by Bayesian inference.
- The assessment of the structural safety (i.e. reliability analysis) based on the inferred parameter distributions to serve decision making.

1.4 Scope and limitations

In this thesis solely a (physical) model driven approach to SHM and system identification is considered. This study uses two sheet pile wall structures modelled in Plaxis 2D and Plaxis 3D (*Brinkgreve, Kumarswamy, Swolfs, Zampich, & Manoj, 2019*) respectively to explore the research questions given above. In case of the 2D sheet pile wall model, the considered number of dimensions of the probabilistic model is up to eight, while some crucial parameters are assumed to be known, i.e. deterministic (e.g. external loading). In this case, synthetic observations are used and generated using Plaxis. Model uncertainty and measurement uncertainty are both assumed to follow a multivariable normal distribution with correlation taken into consideration (model uncertainty is modelled as a Gaussian process regression). It is important to note that in this case, we use location based data rather than time series data, which might not be representative of realistic structural behavior over time. The measurement and model uncertainties are assumed to be fixed, i.e. not updated during the Bayesian inference. Then in the 3D model, which is calibrated to a full-scale experiment, the correlation between model uncertainties of different sensors is investigated. Bayesian inference is not performed as it takes too much time (i.e. more than one hour) to run a single Plaxis analysis.

Even though this study is related to hydraulic structures, the method and the core idea of combining monitoring data, Bayesian inference and reliability analysis can be easily applied to other engineering structures or even other fields as well.

1.5 Contributions

Since the domain of SHM is an active and rapidly evolving research area with contributions of researchers from all over the world, this section aims to mark briefly the main contributions of this work. They are harmonized with the research questions and the literature review. Detailed support for the novelty of these contributions can be found in Chapter 2 in the literature review.

- Combination of system identification and structural reliability for hydraulic structures, first time in the open literature.
- Analysis of the impact of construction stages on SHM of hydraulic structures.
- Application of nested sampling for SHM based inference and reliability analysis; to my knowledge the first time in the open literature.
- Proposal of a novel surrogating approach (active learning Gaussian process regression).
- Proposal for a general mathematical model that encompasses measurement uncertainties, physical model uncertainties, and surrogate model uncertainties. The quantification and inclusion of the latter in the system identification of hydraulic structures is believed to be the first time in the open literature.
- Other implementation related contributions: implementation of a custom, Python-based program with the following features: programmatic control of Plaxis, inference, reliability assessment (reference to Annex A where all the particular packages are listed, given for the various tasks).

1.6 Thesis structure

This thesis has five chapters. First, an introduction including the research motivation and research goals is presented in chapter 1. Chapter 2 reviews the relevant literature on SHM with a particular focus on the domains of hydraulic structures and geotechnical structures. Furthermore, the contributions of this thesis are highlighted. Next, the adopted methodology is elaborated in chapter 3, including the handling of uncertainty, the essence of Bayesian inference, the construction of surrogate models, and the principals of a reliability-based assessment. With this, the ingredients to answer the research questions are introduced. Chapter 4 presents the analysis to two sheet pile wall systems (namely 2D and 3D model) that are deemed to be complex enough to represent real life structures, which demonstrates the feasibility and utility of the proposed methodology, and to answer the research questions. Finally, the conclusions and recommendations are given along with an outlook for future work.

2 Literature review

2.1 Overview

As a preparation to answer the research questions, the relevant literature is studied to survey the current state-of-the-art and to find potential knowledge gaps with focus on hydraulic structures (and particular subtopics are treated in more details in chapter 3). Firstly, general monitoring systems in SHM are reviewed. Then some methods of system identification are discussed with particular focus on Bayesian approaches as it is a sound way to account for uncertainties. Different research studies are presented and their contributions and limitations are discussed. Furthermore, attention is paid to the implementation of several Bayesian computational methods, as well as the use of surrogate models in the civil engineering field. Finally, literature regarding the assessment of structures in the support of decision making is studied.

2.2 Structural health monitoring – Data collection

2.2.1 Definition

During the service life of a civil structure, it may be subjected to natural actions (such as earthquakes, flood, and metal oxidation) and severe accidental impacts due to human behavior (e.g. collision), which can lead to a considerable damage and/or a reduction of its life time. It is crucial to detect such damage as early as possible and evaluate the reliability level of the structure. Besides, it is also crucial to learn the actual conditions of a structure to reduce uncertainty of structural properties (e.g. soil parameters in a sheet pile wall system). One approach to achieve these goals is via structural health monitoring (SHM) that is intensively researched for decades and still an active area with many open challenges. There are generally two components in SHM. One is monitoring, which is the process of collecting information about a structure in a systematic and often fully automated fashion, e.g. accelerometers installed on a bridge, continuously measuring and logging data into a database. The information collection in monitoring refers solely to the process of measuring and logging data that is measured by sensors, i.e. there is no or little post-processing of the data involved. This potential little post-processing is only considered at the sensor level and the structure itself will not be

involved. The second component of SHM is the post-processing of the data that requires knowledge of the structure. An example of that is system identification, which will be further discussed in section 2.3. The explanation of SHM above could also be found in (*Balageas, 2006*), “Structural Health Monitoring (SHM) aims to give, at every moment during the life of a structure, a diagnosis of the “state” of the constituent materials, of the different parts, and of the full assembly of these parts constituting the structure as a whole”.

2.2.2 Examples of application

A complete and mature SHM system usually aims at continuously, remotely, and automatedly monitoring the performance of a structure by using different kinds of sensors to collect responses. This is a solid foundation to acquire and process data and then to detect the potential damage or learn the current condition of structures. And due to the boom of sensor technology, it is now possible to construct complex but efficient monitoring system in different structures. For example, *Dascotte, Strobbe, and Tygesen (2013)* described a wind and structural health monitoring system (WASHMS) that is deployed by the Hong Kong Highways Department to monitor the structural performance of the Stonecutters Bridge and to evaluate possible defects. *H. S. Park, Lee, Adeli, and Lee (2007)* proposed to include terrestrial laser scanning (TLS) in the monitoring system to overcome the difficulties of in situ instrumentation of sensors on large structures (e.g. high-rise buildings). Multiple sensors like inclinometers and Shaped Accel Arrays (SAAs) were installed on the retaining walls of the launch shaft at Aughooose to monitor its behavior after construction in (*Baily, Kieran, Goggins, & McCabe, 2014*). Furthermore, *Fabbrocino, Laorenza, Rainieri, and Santucci de Magistris (2009)* implemented a monitoring system, “Casa dello Studente” to monitor the dynamic behavior of a flexible retaining wall under seismic load. However, considering the efficiency of monitoring system, guidance is seldom available regarding how many and what type sensors should be installed, and where to install them.

2.3 Structural health monitoring – Evaluation

2.3.1 System identification

Once the monitoring system is built and data is collected, the next step in the process of evaluating the “health” of a structure is system identification (*Collins, Hart, Hasselman, & Kennedy, 1974*). This widely adopted approach focuses on the estimation of structural properties from responses collected by sensors installed on a structure. For instance, *Dragos*

and Smarsly (2015) identified the potential loosening of the plate-to-column connections of a four-story frame structure by a comparison between the initial state of a finite element (FE) model and a simulated damaged state. In the hydraulic and geotechnical context, system identification is also broadly used to identify soil parameters. For example, *Levasseur et al. (2008)* adopted the least square method to identify the dilatancy angle and the shear modulus in a hypothetical sheet pile wall system. *Tang and Kung (2009)* proposed an efficient way, nonlinear optimization technique (NOT) to identify soil parameters like elastic modulus and Poisson's ratio in case with deep excavation.

Generally, the aforementioned applications of system identification use the data collected by sensors deterministically, which means that they neglect the potential uncertainty when moving from reality to physical model and from real structural responses to data collected by sensors. However, uncertainties are always present and often substantial in case of civil engineering structures (*JCSS, 2000*), particularly geotechnical structures (*Phoon & Tang, 2019*). Just consider the immensely expensive and hence scarce or often non-existent full-scale experiments which could be used to validate models and reduce modelling uncertainty. The Bayesian paradigm offers a framework to quantify and propagate uncertainties. Multiple researchers have explored its potential, for example, *Kelly and Huang (2015)* adopted Bayesian inference to identify soil parameters like unit weight in a synthetic one dimensional consolidation soil model. Their analysis includes measurement error. *Miro, Konig, Hartmann, and Schanz (2015)* used Bayesian inference with a lumped uncertainty (measurement and statistical uncertainty) in the system identification of a shallow tunnel. *L. L. Zhang et al. (2010)* inferred the cohesion coefficient and the friction angle in a hypothetical slop failure case, considering an additive model uncertainty. *L. Wang, Hwang, Luo, Juang, and Xiao (2013)* included measurement uncertainty in the identification of the slope failure on the NO.3 freeway in Taiwan. *X. Y. Li, Zhang, and Jiang (2016)* and *Ering and Babu (2016)* considered both measurement uncertainty and model uncertainty to infer soil properties in the excavation of a 530m high rock slop and in a rainfall-induced landslide case respectively. Finally, *L. Zhang et al. (2018)* included only measurement uncertainty and used multiple types of observations (e.g. displacement, ground water level, water pressure, etc.) to conduct Bayesian inference in a soil slope under rainfall infiltration. More details of these studies are summarized in Table 2.1.

Though the applications are different, the common limitation of these studies is that the considered dimensionality is quite low, i.e. the considered parameters are relatively small in number (i.e. on average four). In reality, however, more (soil) parameters could be influential.

Another limitation is that the effect of construction stages on system identification is rarely (not) investigated, as most of these studies focus on existing structures. And in their work, uncertainties are usually not explicitly explained (i.e. lumped uncertainties are used), or they are not complete (i.e. only measurement uncertainty or model uncertainty is considered).

2.3.2 Bayesian computational methods

Once the system identification problem is formulated as a Bayesian inference problem, we need a computational method to solve it. For low dimensional problems (i.e. when no more than 3 parameters are to be inferred), the brute force/straightforward numerical integration method can be used and there is no need for e.g. sophisticated sampling based methods. However, numerical integration becomes practically intractable when moving to higher dimensionality (arguments supporting this claim can be found in section 3.4). Therefore, other computation algorithms are adopted to circumvent the integration. Considering the application in hydraulic structures, these algorithms often belong to the group of Markov Chain Monte Carlo (MCMC), among which Metropolis-Hastings (MH) MCMC is particularly preferred. For example, both *L. L. Zhang et al. (2010)* and *L. Wang, Hwang, Luo, Juang, and Xiao (2013)* adopt Metropolis-Hastings (MH) MCMC to replace the calculation of integrals.

However, as summarized in (*MacKay & JC, 2003*), MH MCMC usually has a low acceptance rate and needs a large amount of evaluations. Furthermore, in its vanilla form it is not capable of exploring distributions with separate modes and it has difficulties with degenerate posterior.

2.3.3 Surrogate model

No matter which computational method as described in section 2.3.2 is selected, numerous calculations of the physical model (e.g. FE model) are inevitable, and the computation time could be easily in the order of days, which is never desirable. Hence, often a surrogate model is constructed as a cheap-to-evaluate replacement of the computationally demanding model. Some research studies demonstrate the feasibility of surrogate models in system identification within the hydraulic structures context. For example, *D. Park and Park (2017)* used a first order polynomial model to simulate the ground deformation to identify the deformation modulus and stress ratio of soil in a railway tunnel case. Both *Miro et al. (2015)* and *L. L. Zhang et al. (2010)* used a quadratic polynomial model to replace the costly outputs of FE software. Besides, *Xu and Low (2006)* used a second-order polynomial without intersection terms to surrogate the factor of safety in the assessment of stability of embankments.

Yet with reference to *Forrester, Sobester, and Keane (2008)*, polynomials models (also called response surface models) are not suitable for the highly nonlinear and multidimensional cases that most engineering problems belong to, and once constructed, they provide very limited guidance about where we could continue the sampling process. And they often cannot be scaled to higher dimensional cases while maintaining a reasonable computational time. Therefore, a more flexible global surrogate model which is general and capable of tackling realistic engineering problems should be explored to overcome such limitations.

2.4 Assessment methods

After the system identification is finished, there are typically two ways to decide whether a structure is safe or not as summarized by *Jonkman, Steenbergen, Morales-Napoles, Vrouwenvelder, and Vrijling (2015)*. One is referred as semi-probabilistic method, which for example, the Partial Factor Design, relies on partial factors of inputs and is computationally cheap (e.g. could even be performed by a calculator). For instance, *Ivanov (2013)* adopted the semi-probabilistic method to assess the hull girder section modulus. And *Sykora, Holicky, and Markova (2013)* used the same way in verification of reinforced concrete bridges. The second way to assess the safety of a structure is the reliability-based method (also referred as probabilistic approach), such as Monte Carlo simulation and first order reliability method (FORM), which compute the failure probability of a structure. For example, *Schweckendiek (2006)* used FORM to calculate the structural failure probability (e.g. failure of sheet pile wall and anchor) and used directional adaptive response surface sampling (DARS) to assess the soil failure in a sheet pile system. The whole system failure is then considered by method ‘Hohenbichler’ proposed by *Hohenbichler and Rackwitz (1982)*. *Wolters, Bakker, and De Gijt (2013)* adopted FORM to perform reliability assessment for both soil failure and structural failure. *Teixeira et al. (2015)* used FORM to calculate the failure probability of soil and directional sampling for the structural failure. *Beck and Au (2002)* updated the reliability of a moment-resisting frame by Monte Carlo simulation. *Schweckendiek and Kanning (2016)* combined Bayesian reliability updating method and Monte Carlo simulation to analyze the slope stability of dikes.

Nevertheless, semi-probabilistic approaches trade quick calculation for flexibility, and they are only applicable to cases to which they are calibrated using reliability methods. . And in the literature where a reliability-based method is used, the distributions of input parameters are assumed to be known (e.g. normal distribution with a given mean and standard deviation),

seldom of them performs structural assessment using the results of SHM based system identification.

Table 2.1: Overview of previous works in system identification of hydraulic and geotechnical structures (the variables and acronyms are explained in Table 2.2). Studies with probabilistic system identifications where the uncertainty in the estimated parameters is quantified are highlighted with italic font.

Reference	~ Probabilistic model					~Mechanical model			Surrogate model		Decision making support
	Inferred parameters	Dimensionality	Parameter estimation	Uncertainty	Likelihood function	Measured parameters	Analysed case(s)	Physical model	Model type	Model fitting	
<i>(Arai, Ohta, & Kojima, 1987)</i>	G, q, k	3	LS	No (deterministic)	NA	Displacements (translation) over time	Hypothetical and real	2D LFEA & 2D NLFEA	NA	NA	NA
<i>(Honjo, Wen-Tsung, & Guha, 1994)</i>	E, k	Up to 14	<i>Extended Bayesian method; (optimization to find the smallest AIC)</i>	<i>Measurement noise</i>	<i>MND without correlation</i>	<i>Horizontal and vertical displacements</i>	<i>3 m high control embankment</i>	<i>2D LFEA</i>	NA	NA	NA
<i>(Ledesma, Gens, & Alonso, 1996)</i>	E, K_0	2	ML	<i>Measurement error</i>	<i>MND with correlation</i>	<i>Horizontal and vertical displacements</i>	<i>Hypothetical</i>	<i>2D LFEA</i>	NA	NA	NA
<i>(Gens, Ledesma, & Alonso, 1996)</i>	$3 \times E, K_0$	4	<i>ML and LS</i>	<i>Measurement error</i>	<i>MND with correlation</i>	<i>Vertical, horizontal displacements</i>	<i>Real (tunnel section)</i>	<i>2D LFEA</i>	NA	NA	NA
<i>(J. H. Deng & Lee, 2001)</i>	$6 \times E$	6	LS with GA	No (deterministic)	NA	Displacements	Profile 17–17 of the permanent ship lock slope at the Three Gorges Project	3D FEA linear and nonlinear as well	Back-propagation neural network	NA	NA
<i>(Levasseur, Malécot, Boulon, & Flavigny, 2007)</i>	G, φ, K_0 or G, φ, ψ	3	LS	No (deterministic)	NA	Pressure meter curve and the horizontal displacements	Sheet pile wall (hypothetical based on a real case) and pressure meter test	2D NLFEA	NA	NA	NA

(Tang & Kung, 2009)	$G/\sigma_v, E/G, E_{ur}/G, R_i, v$	5	LS (NOT)	No (deterministic)	NA	Displacements (translation), wall deflection	Deep excavations (hypothetical and real cases)	NLFEA	NA	NA	NA
(L. L. Zhang et al., 2010)	$c, 2 \times \varphi$	3	Bayesian, Metropolis-Hastings	Additive uncertainty	MND without correlation	Observation of slope failure, indicated by safety factor	Hypothetical slope failure; Morgenstern-Price method	NLFEA	Quadratic polynomial response surface	NA	NA
(J. Zhang, Tang Wilson, & Zhang, 2010)	c, φ, K_0 or $3 \times G, \varphi$	3, 4	MAP estimation	Lumped	MND without correlation	Observation of slope failure, indicated by safety factor	Slope stability	Analytical model function	NA	NA	NA
(Moreira et al., 2013)	c, φ, E, K_0 (not all together)	3	LS (GA, ES)	No (deterministic)	NA	Up to 4 measurements	Synthetic case with tunnel excavation	3D NLFEA	NA	NA	NA
(L. Wang, Hwang, Luo, Juang, & Xiao, 2013)	φ, T	2	Bayesian, MCMC	Measurement uncertainty	MND without correlation	Observation of slope failure, indicated by safety factor	Slope failure at a site on Free-way No. 3 in northern Taiwan	Analytical model function	NA	NA	NA
(Kelly & Huang, 2015)	m, H, Y, c_v	4	Bayesian, DRAM	Measurement error	MND without correlation	Settlements and excess pore pressure	Synthetic example one-dimensional consolidation soil model	Analytical model function	NA	NA	NA
(Miro et al., 2015)	$E_{ur}, E_{oed}, \varphi, \gamma_{im}$	4	Bayesian, Metropolis-Hastings	Lumped uncertainty	MND without correlation	Settlements and horizontal displacements	Shallow tunnel	3D NLFEA	Quadratic polynomial response surface	NA	NA
(Chen, Zhao, Ru, & Li, 2015)	$4 \times E, 2 \times K_0$	6	MAP (does not produce)	No (deterministic)	MND with correlation	Displacement	Slope stability of a permanent ship lock within	FLAC	Support vector machine	NA	NA

			uncertainty estimates)			the Three Gorges project					
(X. Y. Li et al., 2016)	$6 \times E, v$	7	Bayesian MCMC (however no uncertainty ranges are reported, only mean values)	Measurement uncertainty, model uncertainty	MND without correlation	Surface displacement	Excavation of a 530 m high rock slope	FLAC 3D	Second order polynomial function	NA	NA
(Ering & Babu, 2016)	c, ϕ, M_s	3	Bayesian (conjugate posterior)	measurement uncertainty, model uncertainty	MND without correlation	Slope behavior	Rainfall-induced landslide	FLAC	NA	NA	NA
(D. Park & Park, 2017)	E, K_0	2	Optimization using successive response surfaces	No (deterministic)	NA	Displacements	Railway tunnel	Analytical model function	First order polynomial with mixed terms	NA	NA
(L. Zhang et al., 2018)	E, k	2	Bayesian, DREAM	Measurement uncertainty	MND without correlation	Displacements, ground water level, pore water pressures, water contents, etc.	Soil slope under rainfall infiltration	2D NLFEA in ABAQUS	NA	NA	NA
This work	E, ϕ, γ, c	Up to 8	MultiNest (nested sampling)	Measurement and model uncertainty (including physical and surrogate model uncertainty)	MND with correlation	Strain, displacement (up to 24 sensors considered)	Synthetic case, Single anchored sheet pile wall	Single anchored sheet pile wall, 2D NLFEA	Gaussian process regression	Adaptive fitting strategy	Prior and posterior reliability analyses

Table 2.2: Explanation of symbols used in Table 2.1.

Symbols	Description	Symbols	Description
q	initial tangent modulus of rigidity	γ_{un}	Unsaturated unit weight
E	Young's modulus	c_v	coefficient of vertical consolidation
E_{ur}	Unloading reloading stiffness	R_f	Failure ratio
E_{oed}	Tangent stiffness for primary oedometer loading	ν	Poisson's ratio
k	Coefficient of permeability	σ_v	Consolidation pressure
K_0	The ratio of horizontal to vertical in situ total stress	MND	Multivariate normal distribution
G	Shear modulus	ML	Maximum likelihood method
ϕ	Friction angle	LS	Least square method
ψ	Dilatancy angle	DRAM	Delayed rejection adaptive Metropolis
c	Cohesion coefficient	DREAM	Differential Evolution Adaptive Metropolis algorithm
T	Anchor force	FLAC	Fast Lagrangian analysis of continua
m	coefficient of volume compressibility	NLFEA	Nonlinear finite element analysis
σ_c	Uniaxial compressive stress	MAP	Maximum a posterior estimation
ρ	Coefficient of geological strength index	GA	Genetical algorithm
M_s	matric suction		

3 Methods and tools

3.1 Overview

This chapter elaborates the adopted methods and tools. To start with and as a preparation of Bayesian inference, a surrogate model needs to be constructed to replace the computationally expensive FE model, which is used as a convenient tool to capture the mechanical behavior of structures. Next, the details of the Bayesian inference are presented: how prior knowledge and measured data are combined while accounting for uncertainties in order to identify system parameters. After that, the performance of inference could be quantitatively evaluated and the reliability assessment which plays a key role in decision making is outlined.

3.2 Surrogate models

Bayesian inference requires a large number ($> 10^3$ - 10^4) of likelihood function evaluations (see section 3.4.1 for further details), each of which requires an evaluation of the physical model. One typical example of physical model is the finite element (FE) model, which uses FE method to numerically solve differential equations that are commonly used to formulate the mathematical model of real structure. Even though FE model is widely used in engineering to model complex mechanical system, each evaluation time of it can vary from minutes to days depending on its scale. In our case considering the large number of evaluations, the computational time is a challenging issue and hence a surrogate model is usually adopted as a cheap-to-evaluate replacement of computationally demanding model.

Surrogate modeling for low-dimensional (≤ 3) cases can be done with at a relatively low computational time by the naive approach of using a uniformly spaced grid. However, with increasing dimensionality the required number of points increases exponentially, hence rendering this naive approach computationally intractable. In these cases adaptive surrogate modeling approaches can be used to reduce the computational time. These approaches are generating new points for fitting the surrogate model in a sequential manner, while the new points are placed at locations where the expected improvement of the surrogate model is large.

An advantage of the naive approach is that the same, single set of analyses can be used to obtain arbitrary number of surrogate models, e.g. horizontal displacement in point A and point B. This

advantage might be lost when the adaptive approach is used as that is specific to one surrogate model, i.e. (to some extent) separate adaptive surrogate model building is needed for point A and point B.

In this study the more involved, adaptive approach is used as the considered dimensionality could be up to eight. Several type of surrogate models and tools are considered and listed in Table 3.1.

With reference to *Owen, Challenor, Menon, and Bennani (2017)*, Gaussian process regression has several advantages in our work and is chosen to build surrogate models. For example, it is suitable for a nonlinear model which most engineering problems are. Moreover, it provides estimation of prediction variance, which allows for an adaptive fitting that is driven by the largest variance (see section 3.2.4). Besides, it is possible to introduce noise by Gaussian process to avoid overfitting, which will be discussed in section 3.2.8. The introduction and derivation of this section are mainly from *Williams and Rasmussen (2006)* and *Forrester et al. (2008)*.

Table 3.1: Overview of surrogating tools.

Name	Language	Function type	Adaptive modeling	Comments
<i>pyKriging</i>	Python	Ordinary kriging	Yes (MSE and EI)	Not maintained anymore
<i>pySOT</i>	Python	GPR, MARS, polynomial, etc.	Yes	For surrogate based optimization, pure surrogating seems to be not available (adaptive sampling is probably driven by the optimization but we need a good fit over the entire support).
<i>SMT</i>	Python	Kriging, polynomial, KPLS, etc.	No	
<i>scikit-learn</i>	Python	GPR, SVR, ridge regression, etc.	No	General machine learning package
<i>SUMO</i>	Matlab	Radial basis function, polynomial, kriging	Yes	
<i>DACE</i>	Matlab	Ordinary kriging	No	Only for fitting kriging models
<i>ooDACE</i>	Matlab	Kriging	No	Object oriented version of DACE

MSE – mean squared error

EI – expected improvement

GPR – Gaussian process regression

SVR – support vector machine

MARS - Multivariate Adaptive Regression Splines

KPLS – Kriging Partial Least Squares

3.2.1 Gaussian process

Williams and Rasmussen (2006) gives a formal definition of Gaussian process (GP), “A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution”. This joint distribution, denoted as $f(\mathbf{x})$, is determined by a mean function $m(\mathbf{x})$ and a covariance function $k(\mathbf{x}, \mathbf{x}')$:

$$m(\mathbf{x}) = E[f(\mathbf{x})] \quad (3.1)$$

$$k(\mathbf{x}, \mathbf{x}') = E[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \quad (3.2)$$

Where \mathbf{x} and \mathbf{x}' indicate any two values (or vectors) of variables.

And this Gaussian process will be written as:

$$f(\mathbf{x}) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (3.3)$$

Since a GP is a collection of random variables, it requires a property of consistency which is automatically met if the elements of the covariance matrix are obtained from the same covariance function. The consistency means that if a certain GP is defined, then any number of samples from this Gaussian process will be a multivariate normal distribution. For example, if a set of samples, $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)$ are drawn from the same GP, then

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \sim N(m(\mathbf{x})_{n \times 1}, \mathbf{K}_{n \times n}) \quad (3.4)$$

Where each entry in $m(\mathbf{x})_{n \times 1}$ is equal to μ , while $\mathbf{K}_{n \times n}$ is the correlation matrix in which each entry is calculated by covariance function $k(\mathbf{x}, \mathbf{x}')$. Equivalently, the sample \mathbf{y} are drawn from a normal distribution with mean of μ and standard deviation σ , and then they are assigned correlation by covariance function $k(\mathbf{x}, \mathbf{x}')$. This property explains how we construct a Gaussian process regression (GPR) model: given a set of samples $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ and the corresponding outputs (i.e. function values) y_1, y_2, \dots, y_n , they could be correlated by the covariance function and reconstructed to approximate the real corresponding GP. Then prediction at a new sampling point could be made by considering the correlation between the new point and the given points. This assumption implies that we expect that the to-be-

surrogated function (i.e. an engineering function) will behave smoothly and continuously . And how the surrogated function behaves largely depends on the choice of covariance function, which will be detailed in section 3.2.1.1.

3.2.1.1 Kernel

The covariance function $k(\mathbf{x}, \mathbf{x}')$, which is called the kernel of GP, indicates the dependence or correlation between two points \mathbf{x}, \mathbf{x}' . The parameters used to define a kernel function are called hyperparameters. The kernel could be classified into stationary and non-stationary types. The former means that the value of kernel function only depends on the relative distance between two points, and the closer the two points are, the more correlated and similar they are, and thus the closer their corresponding function values are. In contrast, the non-stationary kernel also considers the absolute location of the given points. In addition, kernels can also be classified into anisotropic and isotropic types. The former means that the kernel function behaves differently, or has different hyperparameters along different dimensions/axes, while the latter refers to a kernel which has the same hyperparameters in all directions (the terminology is equivalent to the terminology used in structural engineering for materials). In our work, only stationary kernels are considered. Several common kernels taken from [Duvenaud \(2014\)](#) are shown in Figure 3.1, of which two, namely the white kernel and radial basis function (RBF) kernel are detailed in the next two sections.

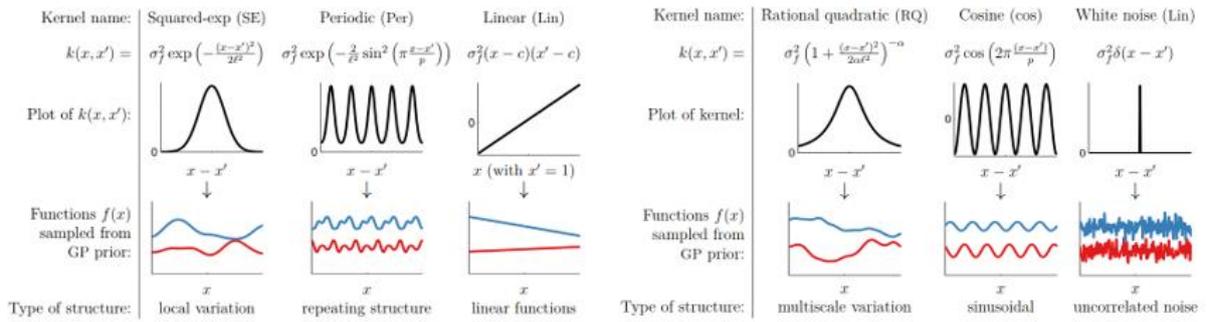


Figure 3.1: Kernels in Gaussian process regression [[Duvenaud \(2014\)](#)].

3.2.1.2 White kernel

This kernel can be used to introduce noise to the GPR model and can be defined by the following equation:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} \alpha & \text{if } \mathbf{x}_i = \mathbf{x}_j \\ 0 & \text{otherwise} \end{cases} \quad (3.5)$$

Where

x_i, x_j the i^{th} and j^{th} data points;

α noise level.

3.2.1.3 Radial basis function (squared exponential) kernel (RBF)

Isotropic case:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left(-\frac{(\mathbf{x}_i - \mathbf{x}_j)^2}{2\ell^2}\right) \quad (3.6)$$

Anisotropic case:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left(-\sum_{k=1}^d \frac{(\mathbf{x}_i^{(k)} - \mathbf{x}_j^{(k)})^2}{2(\mathbf{L}^{(k)})^2}\right) \quad (3.7)$$

Where

d dimensionality (i.e. number of random variables);

ℓ length scale, is a scale for isotropic kernel, indicates the correlation between two sampling points;

\mathbf{L} length scale, is a vector of length d for anisotropic kernel;

σ_f^2 signal variance or function variance.

3.2.2 Model construction of Gaussian process regression

As explained previously, the construction of a Gaussian process regression (GPR) model aims at finding the approximation of corresponding real GP based on the given data and a selected kernel. There generally two steps needed. Firstly, a number of initial sampling points are selected to build a basic “framework” of the surrogate model. Then more sampling points are added adaptively to improve the accuracy.

3.2.2.1 Initial design

To construct a surrogate of the simulator, which refers to the to-be-surrogated function/model from herein, the selection of initial sampling points considering their position and number is crucial. The former could be done based on for example, Latin Hypercube (LH) scheme, while the latter highly depends on the simulator and hence general recommendations of it cannot be

made. However, for expensive simulators it is advised to start with a minimal number of initial points and add additional ones (the strategy for adding further points is explained in section 3.2.4). *Razavi, Tolson, and Burn (2012)* has summarized some recommendations about the number of initial points for different situations which are given in Table 3.2.

Table 3.2: Overview of selected number of initial points.

	Comments
$p = 10dim$	Kriging and RBFs
$p = 2^{dim}$	Samples the corner of variable space
$p = 2(dim+1)$	Kriging and RBFs
$p = \max[2(dim+1), 0.1n]$	If computational budget is permitted, it could achieve better global accuracy with more sampling points.
$p = 0.35n$	An upper bound for deceptive and multimodal simulators

p : approximate number of initial points

dim : dimensionality

n : total number of simulator evaluations, accounts for computational budget.

3.2.2.2 Hyperparameters estimation of kernel

Given an initial set of sampling points, a proper kernel is required to connect these samples and make predictions. Even though there are many kernels available and some of them are even capable to capture periodicity in the output, such a behavior is not expected in our context. After experimenting with different options we settled with the following options for the surrogate modeling:

A composite kernel which is the summation of an anisotropic RBF kernel and a white kernel.

The former component ensures the smoothness of GPR model, while the latter can capture potential numerical noise. The noise part is crucial to avoid numerical error and overfitting. Then the last step to finish the construction of the GPR model is to estimate the hyperparameters from data based on the Maximum likelihood method.

Suppose there are a set of initial sample data, $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}^T$ and the corresponding output vector $\mathbf{y} = \{y_1, y_2, \dots, y_n\}^T$ generated from a GP. Based on the definition, the output \mathbf{y} is

equivalently the observation of the GP with a mean μ and standard deviation σ . Then the probability of the occurrence of these \mathbf{y} can be expressed in a likelihood function:

$$L(\mathbf{y} | \mu, \sigma) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2}} \quad (3.8)$$

The to-be-estimated hyperparameters should maximize the likelihood to assure that the given output \mathbf{y} is most likely to be drawn. But this likelihood function is not complete yet as the correlation between samples is not included. This correlation is represented by kernel function as introduced previously, and if correlation is considered, the likelihood function of \mathbf{y} could be rewritten as:

$$L(\mathbf{y} | \mu, \sigma, \boldsymbol{\theta}^{hyper}) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}} |\mathbf{K}|^{\frac{1}{2}}} e^{-\frac{(\mathbf{y} - \mathbf{1}\mu)^T \mathbf{K}^{-1} (\mathbf{y} - \mathbf{1}\mu)}{2\sigma^2}} \quad (3.9)$$

Where $\mathbf{1}$ is a unit vector with length of n , $\boldsymbol{\theta}^{hyper}$ is the hyperparameters in the kernel function embedded in \mathbf{K} , which is the correlation matrix calculated based on the kernel function:

$$\mathbf{K} = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_n, \mathbf{x}_1) & \cdots & k(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix} \quad (3.10)$$

For simplification and numerical reasons (e.g. if the value is very close to zero, it can cause problems with double precision calculations), this likelihood function is often taken the natural logarithm:

$$\ln(L(\mathbf{y} | \mu, \sigma, \boldsymbol{\theta}^{hyper})) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(2\sigma^2) - \frac{n}{2} \ln(|\mathbf{K}|) - \frac{(\mathbf{y} - \mathbf{1}\mu)^T \mathbf{K}^{-1} (\mathbf{y} - \mathbf{1}\mu)}{2\sigma^2} \quad (3.11)$$

The estimation of m and σ , indicated as $\hat{m}, \hat{\sigma}$ could be obtained by taking the derivatives of this equation:

$$\hat{\mu} = \frac{\mathbf{1}^T \mathbf{K}^{-1} \mathbf{y}}{\mathbf{1}^T \mathbf{K}^{-1} \mathbf{1}} \quad (3.12)$$

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{1}\hat{\mu})^T \mathbf{K}^{-1} (\mathbf{y} - \mathbf{1}\hat{\mu})}{n} \quad (3.13)$$

These two result can be substituted back to equation (3.11) and since the aim is to maximize the likelihood, the constant components could be discarded and the result is:

$$\ln(L(\mathbf{y} | \boldsymbol{\theta}^{\text{hyper}})) \simeq -\frac{n}{2} \ln(2\hat{\sigma}^2) - \frac{n}{2} \ln(|\mathbf{K}|) \quad (3.14)$$

Then $\boldsymbol{\theta}^{\text{hyper}}$ could be found by either global search method like genetic algorithm, or by gradient-ascent based optimization method to maximize equation (3.14). This estimation is transformed into an optimization problem with potentially multiple maxima, hence the optimization is performed multiple times using different random initial values each time. For each time, there are $(1 + \text{dim} + 1)$ hyperparameters to be determined where dim is the dimensionality, corresponding to one signal variance (σ_f), dim length scale (l), and one noise level (α). Since each evaluation of the ln-likelihood function requires the inverse of the correlation matrix that is a square matrix with rows equal to the sample points, and also because more sampling points are needed for high dimensional spaces, the computational time will inevitably increase with higher dimensionality.

3.2.3 Prediction by Gaussian process regression

After the GPR model is constructed and the hyperparameters are estimated, prediction can be made straightforwardly. The core idea is to connect the to-be-predicted output \hat{y} (at position $\hat{\mathbf{x}}$) with given outputs (\mathbf{y}) using the kernel function, and then to maximize the likelihood of all outputs (\hat{y} and \mathbf{y}).

Firstly, \hat{y} and \mathbf{y} are merged into a new vector $\tilde{\mathbf{y}} = \{\mathbf{y}, \hat{y}\}^T$, then the correlation between \hat{y} and \mathbf{y} is indicated by:

$$\hat{\mathbf{K}} = \begin{bmatrix} k(\mathbf{x}_1, \hat{\mathbf{x}}) \\ \vdots \\ k(\mathbf{x}_n, \hat{\mathbf{x}}) \end{bmatrix}_{(n \times 1)} \quad (3.15)$$

Now the new correlation matrix is:

$$\tilde{\mathbf{K}} = \begin{bmatrix} \mathbf{K} & \hat{\mathbf{K}} \\ \hat{\mathbf{K}}^T & 1 \end{bmatrix} \quad (3.16)$$

So the \ln -likelihood of all outputs is:

$$\ln(L(\tilde{\mathbf{y}})) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(2\hat{\sigma}^2) - \frac{n}{2} \ln(|\tilde{\mathbf{K}}|) - \frac{(\tilde{\mathbf{y}} - \mathbf{1}\hat{\mu})^T \tilde{\mathbf{K}}^{-1} (\tilde{\mathbf{y}} - \mathbf{1}\hat{\mu})}{2\hat{\sigma}^2} \quad (3.17)$$

And only the last term depends on \hat{y} , hence the \ln -likelihood can be written as:

$$\ln(L(\tilde{\mathbf{y}})) \simeq -\frac{(\tilde{\mathbf{y}} - \mathbf{1}\hat{\mu})^T \tilde{\mathbf{K}}^{-1} (\tilde{\mathbf{y}} - \mathbf{1}\hat{\mu})}{2\hat{\sigma}^2} = -\frac{\begin{pmatrix} \mathbf{y} - \mathbf{1}\hat{\mu} \\ \hat{y} - \hat{\mu} \end{pmatrix}^T \begin{pmatrix} \mathbf{K} & \hat{\mathbf{K}} \\ \hat{\mathbf{K}}^T & 1 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{y} - \mathbf{1}\hat{\mu} \\ \hat{y} - \hat{\mu} \end{pmatrix}}{2\hat{\sigma}^2} \quad (3.18)$$

The maximum of equation (3.18) can be found easily taking derivative to \hat{y} and setting the equation to zero. This will obtain the ultimate prediction of \hat{y} :

$$\hat{y} = \hat{\mu} + \hat{\mathbf{K}}^T \mathbf{K}^{-1} (\mathbf{y} - \mathbf{1}\hat{\mu}) \quad (3.19)$$

Afterwards, the mean squared error (MSE) or variance of this prediction can be computed as:

$$\hat{s}^2 = \hat{\sigma}^2 \left[1 + \alpha - \hat{\mathbf{K}}^T \mathbf{K}^{-1} \hat{\mathbf{K}} + \frac{1 - \mathbf{1}^T \mathbf{K}^{-1} \hat{\mathbf{K}}}{\mathbf{1}^T \mathbf{K}^{-1} \hat{\mathbf{K}}} \right] \quad (3.20)$$

The elaboration of this variance can be found in (Hoyle, 2006) and it is the key in adaptive fitting in next section.

3.2.4 Adaptive fitting of Gaussian process regression model

Generally, a surrogate constructed with initial sampling point is not accurate enough due to the limited number of initial points. It can be improved by adding additional sampling points based on a certain criterion (acquisition function). This process is referred as adaptive fitting or infilling. Forrester et al. (2008) summarizes two general ways to perform infilling. One is exploitation, which focuses on a particular region that is relevant from the application point of view, for example in case of optimization the region of the current optimum, the next infilling

point would be at where the expected improvement is the maximum. This is called local surrogate in this thesis. The other is exploration, which focuses on a general global improvement of the surrogate model and requests the next point where the predicted variance, or MSE is the largest. This is called global surrogate in this work.

Considering the application of local surrogate in Bayesian inference, a few studies are summarized in Table 3.3. They are based on different infilling criteria, of which the core idea is to infill the next point to obtain the largest reduction in differential entropy or KL divergence. This is a much more efficient method as most of the infilling points will be around the peaks of likelihood or posterior.

Table 3.3: A cursory overview of studies using local surrogating.

	simulator	Infilling criterion
<i>(Kandasamy, Schneider, & Póczos, 2017)</i>	likelihood	negative expected divergence (NDE) + exponentiated variance(EV)
<i>(J. Li & Marzouk, 2014)</i>	posterior	KL divergence
<i>(H. Wang & Li, 2018)</i>	posterior	entropy
<i>(Takhtaganov & Müller, 2018)</i>	physical model	expected improvement

As for the global surrogate, the infilling criterion could be used to add the next point where the largest MSE or variance calculated by equation (3.20) occurs. This maximum could be found by global search method, e.g. genetic algorithm (GA) or by gradient based techniques. Global surrogating is a more general problem and not limited to Bayesian inference as the simulator is usually the physical model (e.g. FE model). A comparison is given between global surrogate and local surrogate in Table 3.4. Also considering the advantage of global surrogate that it could be reused in reliability assessment in section 3.5, global surrogate is adopted in this thesis.

Table 3.4: Comparison between global surrogate and local surrogate.

Global surrogate	Local surrogate
Surrogate of physical model is built with paying equal attention to all regions of the support. The total volume of the support is increasing exponentially with increasing dimensions.	Surrogate of physical model, likelihood or posterior is built with paying more attention to regions of the support which are identified to be important for the analyzed case. The total volume of the important regions is expected to be much smaller than the total volume of the support.
<p>Good reusability.</p> <p>Once the surrogate model is constructed, it can be used for:</p> <ul style="list-style-type: none"> • new observations of the same sensors; • new combination of sensors; • different likelihood function (with some limitations), • different priors (with some limitations). 	<p>Not reusable.</p> <p>The surrogate model is built iteratively based on the obtained likelihood or posterior given a specific value of observation. In other words, it focuses on the accuracy only in the concentrated zone in likelihood or posterior.</p> <p>The surrogate model should be reconstructed if:</p> <ul style="list-style-type: none"> • new observations are used • new combination of sensors are used, • the likelihood function changes, • the prior changes.
It takes considerable computation time to achieve the global convergence, and some of the work is redundant (e.g. the extreme value of parameters might never be used).	Computationally efficient as only the most crucial part is surrogated.

3.2.5 Stopping criterion

In this study the following stopping criterion is proposed and used to break the process of adaptive fitting:

$$\sigma_{\max} - noise < \varepsilon_{\sigma} \cdot y_{\text{range}} \quad (3.21)$$

where

- σ_{\max} The maximum of the predicted standard deviation (square root of equation (3.20));
- noise* Square root of noise level in white kernel; it is set to zero if white kernel is not used.
- ε_{σ} A percentage controlling the stopping criterion.
- y_{range} The range of the output of all sampling points (i.e. absolute difference between the maximum and the minimum).

I did not find a general rule to set the value of ε_{σ} as it highly depends on the simulator and the dimensionality. In our work, a value of $\varepsilon_{\sigma} = 0.005$ works quite well to achieve a very low value of relative error (RE), which is defined as:

$$RE = \frac{y_{\text{predict}} - y_{\text{simulator}}}{y_{\text{simulator}}} \quad (3.22)$$

To demonstrate the adequacy of the stopping criterion an example with four dimensions is given. A set of (i.e. 166) initial sampling points are generated in Plaxis, then more sampling points are added based on the previous infilling criterion. After each additional point is added, the GPR model is updated and the value of $\sigma_{\max} - \text{noise}$ is displayed in Figure 3.3. Then 300 testing points are used to calculate the relative difference, of which the maximum is displayed in Figure 3.2. It could be seen from both figures that after $\varepsilon_{\sigma} = 0.005$ is satisfied (i.e. around 300 sampling points), the maximum of RE is stationary.

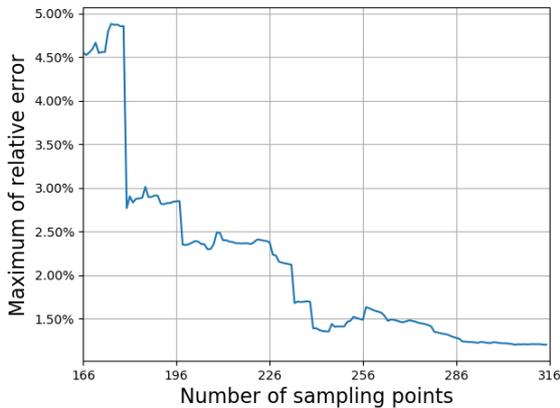


Figure 3.2: Variance of RE with number of sampling points.

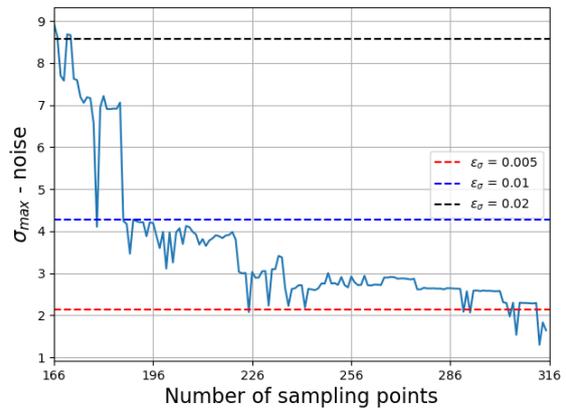


Figure 3.3: Variance of threshold with number of sampling points.

3.2.6 Pseudo code

For the general case of fitting a single surrogate model the pseudo code of the algorithm proposed and used in this study is given in Table 3.5. Firstly, a set of initial sampling points are selected based on LH scheme and the corresponding simulator values are calculated. Then these samples are used to construct a GPR model, in which the hyperparameters are optimized with 10 different initial values. The next step is to add additional sampling point one by one at where the maximum of prediction variance occurs using genetic algorithm. If the square root of this maximum variance excluding the noise component is below the threshold (i.e a certain percentage of the range of all simulator values), the infilling is terminated and the construction of this surrogate model is done.

This process could be repeated for different surrogate models. In our case, the simulator is a FE model, meaning that the previous results can be reused. For example, suppose there is a case where the horizontal displacement at top point and middle point along a sheet pile wall are to be surrogated. When a certain sampling point is given to run the FE software, the horizontal displacement along the whole sheet pile wall can be saved. Then more points are added based on the stopping criterion at the top point and after the GPR model at the top point is converged, all the data including initial and additional sampling points as well as their corresponding response could be used as the initial input for the middle point. This route is reiterated if more surrogate models are required. In practice, this method works well as fewer and fewer infilling points are needed for the sequent surrogates.

Table 3.5: Pseudo code of the proposed adaptive Gaussian process surrogate method.

General pseudo code	Implementation details
generate x_{set} from \mathcal{X} , $\dim(x_{\text{set}}) = (n_{\text{initset}}, k)$	Latin hypercube sampling to select the initial x_{set}
calculate $y_{\text{set}} = \text{Simulator}(x_{\text{set}})$	The simulator could be e.g. FE model
set $\text{convergence} = \text{False}$	
while convergence is False	
calculate $y_{\text{range}} = [\min(y_{\text{set}}), \max(y_{\text{set}})]$	
fit a GPR to $[x_{\text{set}}, y_{\text{set}}]$: <i>GPR</i>	Maximum likelihood estimation with 10 different initial points
find $\sigma_{\text{max}} = \max(\sigma_{\text{pred,GPR}}(x))$ $x \in \mathcal{X}$ and $x_{\sigma_{\text{max}}}$	Global search (genetic algorithm)
if $\sigma_{\text{pred,GPR,max}} - \text{noise} \leq \varepsilon_{\sigma} \cdot y_{\text{range}}$	<i>noise</i> : square root of noise level in white kernel if it is used, otherwise set to zero
set $\text{convergence} = \text{True}$	
else	
calculate $y_{\text{new}} = \text{Simulator}(x_{\sigma_{\text{max}}})$, $x_{\text{new}} = x_{\sigma_{\text{max}}}$	
add to set $x_{\text{set}} = [x_{\text{set}}, x_{\text{new}}]$, $y_{\text{set}} = [y_{\text{set}}, y_{\text{new}}]$	
return <i>GPR</i>	

3.2.7 Illustrative example

In this section, a simple example taken from *Pedregosa et al. (2011)* is shown to illustrate the whole process of global surrogating. Considering the symbolic function: $f(x) = x \sin(x)$, 6 initial sampling points are selected and indicated by red dots shown in Figure 3.4. The RBF kernel is selected and the two hyper-parameters, namely signal variance σ_f and length scale l are estimated using the maximum likelihood method. Since there is considerably large prediction standard deviation at the tail, some additional sampling points are needed. One more sampling point can be added at $x = 10$ where the maximum of standard deviation σ calculated by equation (3.20) occurs. This point is indicated by a black dot in this figure. In this 1D problem (only one variable), this maximum can be found by calculating all the σ along the support, but if more than one variable exists, more advanced methods like GA can perform this work properly. Even though gradient based optimizers are also widely used to find the maximum, they might be trapped in a local maximum considering this σ function might have multiple maximums.

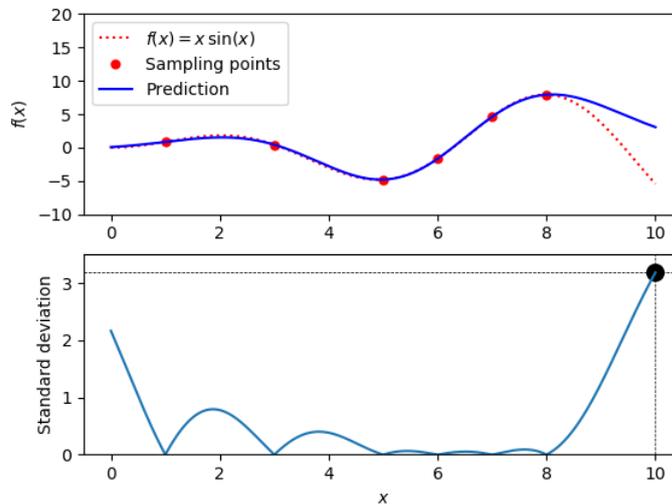


Figure 3.4: Initial sampling points and fit.

After the extra sampling point is added the hyper-parameters of the kernel are re-optimized to maximize the likelihood. Even if the prediction and real function curve match well as displayed in Figure 3.5, there is still large standard deviation at $x = 0$. This is determined by the kernel function, as there are no sampling points near the left support point ($x = 0$). This can be shown more clearly in Figure 3.6 if prediction is performed far from the left support point where larger standard deviation would occur.

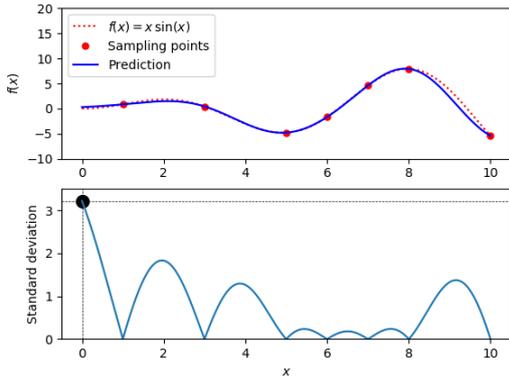


Figure 3.5: Demonstration of predictive error (1).

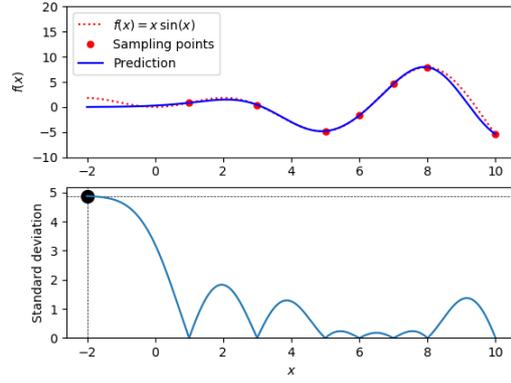


Figure 3.6: Demonstration of predictive error (2).

Since the true simulator function is never known beforehand, it is necessary to add additional points at the boundary of supports or include them as initial sampling points to decrease the standard deviation.. Compared with Figure 3.5, not only do the prediction and real function matches well, but the corresponding standard deviation is also very low if one more point is added at $x = 0$ and shown in Figure 3.7.

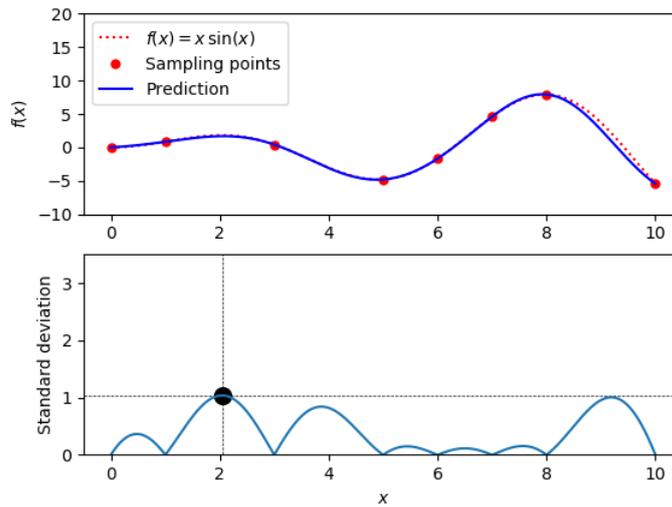


Figure 3.7: Demonstration of predictive error (3).

3.2.8 Influence of noise

The noise added in GPR model plays a crucial role in preventing overfitting. Overfitting can not only be observed from the response surface, but also be checked by the length scale of use RBF kernel. In this section, a simple 2D (i.e. dimensionality is 2) example is given to support this claim.

In this example, the samples are still from Plaxis and the earlier shown steps in Table 3.5 are followed to build surrogate models, which are shown in Figure 3.8 and Figure 3.9. The only

difference between them is that the former uses solely a single RBF kernel while the latter uses an RBF as well as a white kernel. The hyperparameters are indicated above each figure (the number before RBF represents the signal variance).

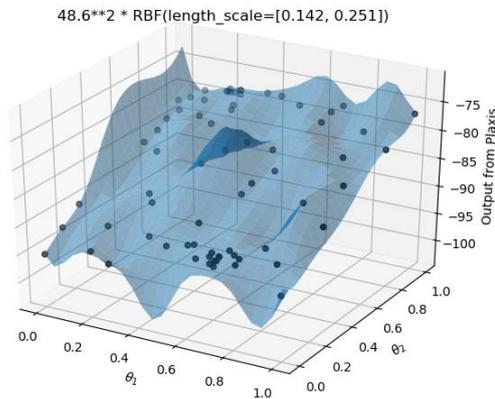


Figure 3.8: Response without noise component.

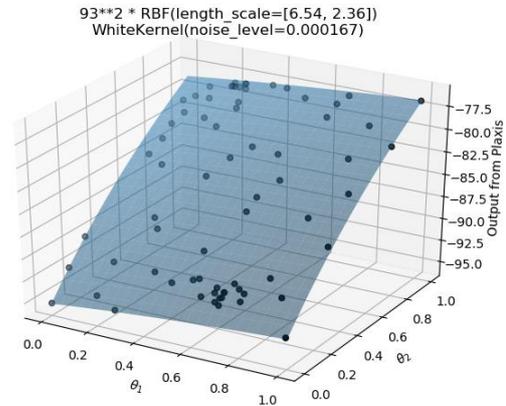


Figure 3.9: Response with noise component.

Intuitively, the one with a noise component on the right seems to be more reasonable as it is much more smooth, that corresponds to our understanding of the underlying physics. The superiority of the GPR model with noise term can be quantitatively demonstrated by calculating the relative difference with 300 test points. The results are shown in Figure 3.10 and Figure 3.11 respectively.

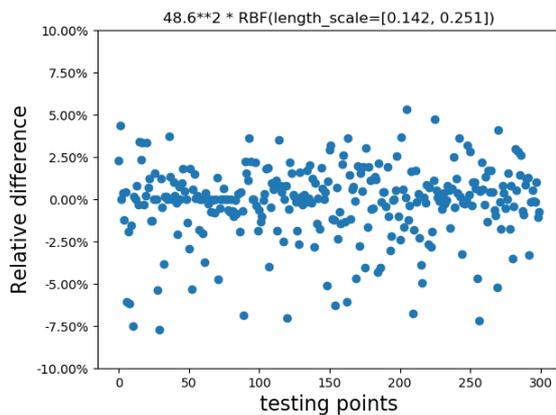


Figure 3.10: RE without noise component.

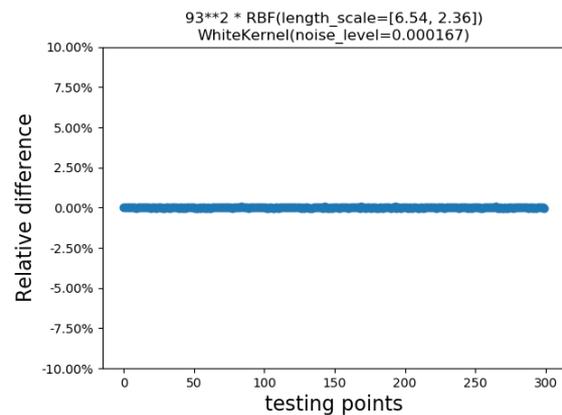


Figure 3.11: RE with noise component

From these two figures, it is clear that the GPR model with noise term does a much better job in prediction due to its considerably low relative error. Even though the magnitude of noise level is negligible compared with outputs (i.e. 0.000167 vs tens), it plays the crucial role to prevent the GPR model going through all sampling points precisely, and thus overfitting could

be avoided. Besides computing the relative difference, occurrence of overfitting could be quickly checked by the length scale of the used RBF. Overfitting usually appears with a very small length scale (e.g. 0.1 ~ 0.2 in the first GPR model), meaning that there is limited correlation between two nearby sampling points, which is inconsistent with the core idea of RBF kernel.

3.3 Uncertainty

3.3.1 Measurement uncertainty

When collecting data from sensors, the observations are inevitably contaminated with measurement uncertainty, which reflects the discrepancy between the measured data and the true/real value. Measurement uncertainty can be caused for example by incorrect reading and/or installation of a measuring instrument. In this study for a particular sensor, we assume a simple additive relationship between the real and measured value:

$$D_{\text{measure}} = d_{\text{real}} + E_{\text{measure}} \quad (3.23)$$

Where

- d_{real} true value of response;
- D_{measure} the value of response measured by the sensor;
- E_{measure} measurement uncertainty (random variable).

To aid the reader the used notation is summarized and explained in Table 3.6.

Throughout this work, multiple sensor types, sensors instances, and load cases will be considered, hence the notation that will be used in later sections is introduced here.

Table 3.6: Explanation of notation.

	Example	Description
Italic capital letter	D	Random variable
Italic smaller case letter	d	Deterministic value (known or unknown)
Bold letter	\mathbf{d}, \mathbf{D}	Vector or matrix

For this case, equation (3.23) can be generalized to:

$$D_{\text{measure},i,j,k} = d_{\text{real},i,j,k} + E_{\text{measure},i,j,k} \quad (3.24)$$

- i i varies with different types of sensors or corresponding responses, e.g. strain, displacement. In this work, same type of response or sensors but of different structural elements are assumed to have different i values, e.g. strain of sheet pile wall and strain of anchor will have different i values.
- j j varies with different positions for a certain type- i response, e.g. the top and middle point along the sheet pile wall when considering its strain.
- k k varies with different load cases (e.g. different construction stages)
- $d_{\text{real},i,j,k}$ true value of the j^{th} type- i response under the k^{th} load case
- $D_{\text{measure},i,j,k}$ the value of the j^{th} type- i response under the k^{th} load case measured by sensors
- $E_{\text{measure},i,j,k}$ measurement uncertainty of j^{th} type- i response under k^{th} load case

In this work, $E_{\text{measure},i,j,k}$ is modeled with a normal distribution with zero mean and standard deviation of $\sigma_{\text{measure},i,j,k}$, described below:

$$E_{\text{measure},i,j,k} \sim N(0, \sigma_{\text{measure},i,j,k}^2) \quad (3.25)$$

It is assumed that each pair of $E_{\text{measure},i,j,k}$ are mutually independent. If multiple sensors are used, the $\mathbf{E}_{\text{measure}}$ will be a multivariate normal distribution (MVN) with zero mean vector and covariance matrix of $\mathbf{\Sigma}_{\text{measure}}$ which is a diagonal matrix with each entry of $\sigma_{\text{measure},i,j,k}^2$:

$$\mathbf{E}_{\text{measure}} \sim MVN(\mathbf{0}, \mathbf{\Sigma}_{\text{measure}}) \quad (3.26)$$

The justification of this additive Gaussian error model is slim and two-fold:

- It is simple and prevalently applied in statistics and other fields.
- We assume that measurement uncertainty is influenced by many processes and their additive effect is observed as a random variable. The central limit theorem supports the usage of a normal distribution in these situations.

3.3.2 Model uncertainty

3.3.2.1 Physical model uncertainty

An accurate physical model (e.g. FE model) is always essential to perform a mechanical analysis of a structure. In practice however, the physical model is always different than the real structure. This difference is stemming from simplifications and assumptions when building FE models. For example, the difference shows up when a sinusoidal function is modeled by a polynomial function, or when some loads are neglected, e.g. influence of temperature on structures. The details of the process that causes the difference is unknown³, hence it is replaced with a probabilistic model, this is commonly called model uncertainty. In this study this relation is assumed to have the following mathematical structure:

$$D_{\text{physical model},i,j,k} = d_{\text{real},i,j,k} + E_{\text{physical model},i,j,k} \quad (3.27)$$

Where

$D_{\text{physical model},i,j,k}$ the value of j^{th} type- i response under k^{th} load case computed by physical model

$E_{\text{physical model},i,j,k}$ physical model uncertainty of j^{th} type- i response under k^{th} load case

In this work, it is assumed that $E_{\text{physical model},i,j,k}$ follows as a normal distribution with mean of zero, and standard deviation of $\sigma_{\text{physical model},i,j,k}$:

$$E_{\text{physical model},i,j,k} \sim N(0, \sigma_{\text{physical model},i,j,k}^2) \quad (3.28)$$

If multiple sensors are used, $\mathbf{E}_{\text{physical model}}$ becomes a multivariate normal distribution with mean vector of zero, and covariance matrix of $\mathbf{\Sigma}_{\text{physical model}}$:

$$\mathbf{E}_{\text{physical model}} \sim MVN(\mathbf{0}, \mathbf{\Sigma}_{\text{physical model}}) \quad (3.29)$$

Consideration of correlation for model uncertainty is crucial as the assumption of independence would lead to unrealistic unacceptable behavior: a model with independent components would tell us that placing more and more sensors at the very same location brings us new information.

³ This includes the case of chosen ignorance as well, i.e. unknown in the selected model universe.

In the covariance matrix, each entry is described as:

$$\Sigma_{\text{physical model}}^{m,n} = \rho_{(i,j,k) \leftrightarrow (i',j',k')} \sigma_{\text{physical model},i,j,k} \sigma_{\text{physical model},i',j',k'} \quad (3.30)$$

$$\rho_{(i,j,k) \leftrightarrow (i',j',k')} = \hat{\rho} e^{-\frac{d_{(i,j,k) \leftrightarrow (i',j',k')}}{l_i}} \quad (3.31)$$

where

$\Sigma_{\text{physical model}}^{m,n}$	m^{th} row and n^{th} column entry in the covariance matrix Σ_{physical}
$\sigma_{\text{physical model},i,j,k}$	physical model uncertainty of related m^{th} sensor in covariance matrix (i.e. j^{th} type- i sensor under k^{th} load case)
$\sigma_{\text{physical},i',j',k'}$	physical model uncertainty of related n^{th} sensor in covariance matrix (i.e. and j'^{th} type- i' sensor under k'^{th} load case)
$\rho_{(i,j,k) \leftrightarrow (i',j',k')}$	correlation coefficient of j^{th} type- i sensor under k^{th} load case and j'^{th} type- i' sensor under k'^{th} load case
$d_{(i,j,k) \leftrightarrow (i',j',k')}$	Euclidean distance between j^{th} type- i sensor under k^{th} load case and j'^{th} type- i' sensor under k'^{th} load case, always a positive value.
l_i	correlation length of type- i sensor
$\hat{\rho}$	An additional term to account for the potential correlation between different type of sensors (i.e. different i values). It is set as 1.0 if $i = i'$, and smaller than 1.0 (e.g. 0.8) if $i \neq i'$

3.3.2.2 Surrogate model uncertainty

When a surrogate model is adopted to replace a physical model, the difference between them should also be considered and included in model uncertainty. This uncertainty is prevalently neglected in the probabilistic system identification literature. In this study this uncertainty is explicitly considered and expressed in an additive form:

$$D_{\text{physical model},i,j,k} = d_{\text{surrogate},i,j,k} + E_{\text{surrogate},i,j,k} \quad (3.32)$$

where

$d_{\text{surrogate},i,j,k}$	the value of j^{th} type- i response under k^{th} load case predicted by the surrogate model
------------------------------	--

$E_{\text{surrogate},i,j,k}$ surrogate model uncertainty of j^{th} type- i response under k^{th} load case

In this work, $E_{\text{surrogate},i,j,k}$ is modeled with a normal distribution with mean of zero, and standard deviation of $\sigma_{\text{surrogate},i,j,k}$:

$$E_{\text{surrogate},i,j,k} \sim N(0, \sigma_{\text{surrogate},i,j,k}^2) \quad (3.33)$$

If the surrogate model is constructed by GPR, $\sigma_{\text{surrogate},i,j,k}$ could be explicitly computed by equation (3.20) and taken the square root. Note the above mathematical model and probabilistic model are not approximations but exact representations and this is how GPR works.

If multiple sensors are used, $\mathbf{E}_{\text{surrogate}}$ becomes a multivariate normal distribution with mean vector of zero, and covariance matrix of $\Sigma_{\text{surrogate}}$ considering correlation:

$$\mathbf{E}_{\text{surrogate}} \sim MVN(\mathbf{0}, \Sigma_{\text{surrogate}}) \quad (3.34)$$

Where each entry (m^{th} row, n^{th} column) in covariance matrix is described as:

$$\Sigma_{\text{surrogate}}^{m,n} = \rho_{(i,j,k) \leftrightarrow (i',j',k')} \sigma_{\text{surrogate},i,j,k} \sigma_{\text{surrogate},i',j',k'} \quad (3.35)$$

The notations above are the same with those in physical model uncertainty.

3.3.3 Combined uncertainty

In this study, the measurement uncertainty and model uncertainty described in section 3.3.1 and 3.3.2 are combined. The combination of all three sources of uncertainty gives the relation between $d_{\text{real},i,j,k}$ and $d_{\text{surrogate},i,j,k}$:

$$d_{\text{real},i,j,k} = d_{\text{surrogate},i,j,k} + E_{\text{measure},i,j,k} + E_{\text{physical model},i,j,k} + E_{\text{surrogate},i,j,k} \quad (3.36)$$

Given the assumptions in each component, the three uncertainties can be combined into a single normal distribution $E_{\text{total},i,j,k}$:

$$E_{\text{total},i,j,k} \sim N(0, \sigma_{\text{total},i,j,k}^2) \quad (3.37)$$

Where

$$E_{\text{total},i,j,k} = E_{\text{measure},i,j,k} + E_{\text{physical model},i,j,k} + E_{\text{surrogate},i,j,k} \quad (3.38)$$

$$\sigma_{\text{total},i,j,k}^2 = \sigma_{\text{measure},i,j,k}^2 + \sigma_{\text{physical model},i,j,k}^2 + \sigma_{\text{surrogate},i,j,k}^2 \quad (3.39)$$

And if multiple sensors are used, $\mathbf{E}_{\text{total}}$ becomes a multivariate normal distribution:

$$\mathbf{E}_{\text{total}} \sim \text{MVN}(\mathbf{0}, \mathbf{\Sigma}_{\text{total}}), \quad (3.40)$$

$$\mathbf{\Sigma}_{\text{total}} = \mathbf{\Sigma}_{\text{measure}} + \mathbf{\Sigma}_{\text{physical model}} + \mathbf{\Sigma}_{\text{surrogate}}. \quad (3.41)$$

The potential correlation is already contained in each component.

3.3.4 Correlation length

The covariance matrix is largely dependent on the selection of correlation length, on which however I did not find any studies to provide guidance. Hence in a later section, the correlation length is derived as the following: if we have access to both d_{real} and $D_{\text{physical model}}$ and their difference is modeled using GPR with RBF kernel, then the length scale of RBF kernel is naturally the correlation length of $E_{\text{physical model}}$ according to its definition in section 3.2.1.3. Even though d_{real} is usually unavailable due to the contamination of E_{measure} , D_{measure} could approximately replace d_{real} to compute correlation length, as E_{measure} is in reality tiny. Since $E_{\text{surrogate}}$ varies not only with sensor type, sensor position and load case, but also with sampling points, meaning that we need a GPR that maps to a vector space (in contrast with the scalar space as in $E_{\text{physical model}}$), it is difficult to determine its correlation length with the same method. Therefore, it is assumed in this work that $E_{\text{physical model}}$ and $E_{\text{surrogate}}$ have the same correlation length for simplicity.

3.4 Bayesian inference

3.4.1 Parameter estimation

Bayesian inference is a statistical inference method in which Bayes' theorem is applied to use data to identify unknown parameters. In Bayesian statistics, all unknown parameters are regarded as random variables and thus have a probability distribution. An advantage of this

thinking is that it is easy to implement further reliability analysis where distribution of parameters is needed. The core idea of Bayesian inference is Bayes' rule, in which a prior knowledge (i.e. probability distribution) of random variables (i.e. refers to soil parameters in this thesis) is considered before any data (i.e. refers to responses of structures) is collected, then data is embedded into the likelihood function and the combination of prior and likelihood leads to a new distribution, the posterior distribution, which contains information of the prior and data.

Bayes' rule for discrete variables:

$$p(\boldsymbol{\theta} | \mathbf{D}) = \frac{p(\mathbf{D} | \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})}{p(\mathbf{D})} = \frac{p(\mathbf{D} | \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})}{\sum_i p(\mathbf{D} | \theta_i) \cdot p(\theta_i)} \quad (3.42)$$

For continuous variables:

$$p(\boldsymbol{\theta} | \mathbf{D}) = \frac{p(\mathbf{D} | \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})}{p(\mathbf{D})} = \frac{p(\mathbf{D} | \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})}{\int_{\theta} p(\mathbf{D} | \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}) \cdot d\boldsymbol{\theta}} \quad (3.43)$$

$$p(\mathbf{D} | \boldsymbol{\theta}) \sim MVN(f(\boldsymbol{\theta}), \boldsymbol{\Sigma}_{\text{total}}) \quad (3.44)$$

where

\mathbf{D}	data (i.e. here refers to observations of responses of structures)
$\boldsymbol{\theta}$	random variables (i.e. soil parameters to be inferred)
$p(\boldsymbol{\theta})$	prior distribution of random variables
$p(\mathbf{D} \boldsymbol{\theta})$	likelihood function, can be any form and equation (3.44) is a particular case to our application.
$p(\mathbf{D})$	a constant known as evidence, obtained by integral or summation over the entire support
$p(\boldsymbol{\theta} \mathbf{D})$	posterior distribution, an update of prior distribution
$f(\boldsymbol{\theta})$	response computed by physical model or surrogate model
$\boldsymbol{\Sigma}_{\text{total}}$	Combined uncertainty described in section 3.3

3.4.2 Overview of computation algorithms

If the number of unknown parameters are no more than three (the dimensionality of problems is no more than three), it is quite straightforward and computationally cheap to perform Bayesian inference or parameter estimation by directly calculating equation (3.43). But for higher dimensional problems (e.g. ten parameters are considered simultaneously), it is not desirable or even not feasible to calculate the integral in the denominator. Therefore, smarter and more advanced methods are usually adopted, which can directly draw samples from the posterior distribution and thus the calculation of evidence can be avoided. Most of these methods belong to the group of Markov Chain Monte Carlo (MCMC) methods, such as Metropolis-Hastings MCMC (MH) in (*MacKay & JC, 2003*), Hamiltonian Monte Carlo (HMC) in (*Neal, 2011*) and Transitional MCMC (TMCMC) in (*Ching & Chen, 2007*). In addition to these, *Feroz, Hobson, and Bridges (2009)* improved nested sampling developed by *Skilling (2006)*, coined as MultiNest, which is also capable of doing Bayesian inference. The core difference among these methods is how they draw the next sampling point given the existing one and their main characteristics are collected in Table 3.7.

Table 3.7: Comparison of different computational algorithms.

	Advantages	Disadvantages
MH	Easy to implement	Low acceptance rate Inefficient for correlated variables Not for multi-modal distributions
HMC	High acceptance rate Less iterations than MH	Needs the calculation of the gradient Difficult to tune step size Much computational time Not for multi-modal distributions
TMCMC	Suitable for multi-modal distributions Versatile: sample from posterior as well as estimate the evidence	Much computational time needed Accuracy decreases with the increase of dimension
MultiNest	Fast Suitable in multi-modal distributions	Difficult to decide the initial number of active sampling points Difficult to implement MCMC diagnose as the number of sampling points is uncertain

3.4.3 Nested sampling

Considering all the aspects of Table 3.7 and based on the comparison of them (e.g. computational time, convergence, etc.) via simple examples, MultiNest is selected to perform the case studies. Since it is based on nested sampling, the description of this method will start with nested sampling.

3.4.3.1 Sampling idea

As introduced by *Skilling (2006)*, in this method, the prior volume is split into a set of elements with equal mass dX :

$$dX = p(\theta)d\theta \quad (3.45)$$

so the total prior mass X could be accumulated from these elements in any order, so we define this equation:

$$X(\lambda) = \int_{P(D|\theta) > \lambda} p(\theta) d\theta \quad (3.46)$$

as the cumulative prior mass which contains all likelihood values larger than λ . The prior mass X will decrease from 1 to 0 with the increase of λ . Then the calculation of evidence is written as:

$$Z = \int_0^1 X(\lambda) d\lambda \quad (3.47)$$

and if the inverse function of $X(\lambda)$ is used, which is positive and monotonically decreasing, the evidence is finally transformed into a one-dimensional integral over the unit space regardless of the original dimensionality:

$$Z = \int_0^1 L(X) dX \quad (3.48)$$

Then for a given sequence of X values, $0 < X_m < X_{m-1} < \dots < X_2 < X_1 < 1$, trapezium rule could be used to approximate the evidence by:

$$Z \approx \sum_{i=1}^m L_i w_i \quad (3.49)$$

where $L_i = L(X_i)$, $w_i = \frac{(X_{i-1} - X_{i+1})}{2}$ represents the weight of each sample.

A picture taken from ([Skilling, 2006](#)) shown in Figure 3.12: could help to indicate this idea:

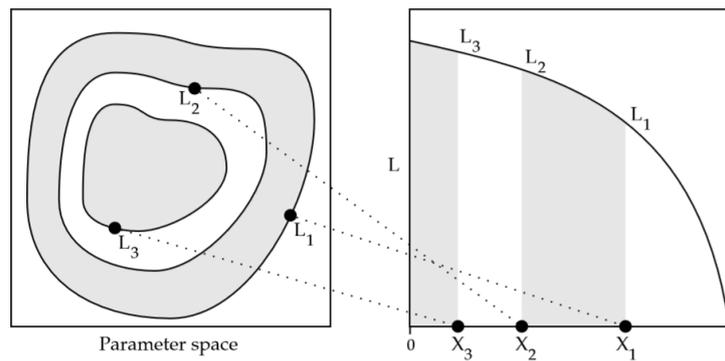


Figure 3.12: Showcase of nested sampling

3.4.3.2 Implementation

The general implementation details of nested sampling are given in the following way:

- N live points or active points are sampled randomly from the full space of prior, and the total evidence is initialized (i.e. $Z = 0, X_0 = 1$);
- At each iteration i , the prior could be estimated as $X_i = e^{-\frac{i}{N}}$. the point with the lowest likelihood (L_i) and its corresponding prior volume (X_i) will be restored;
- The total evidence Z is increased by $L_i \frac{(X_{i-1} - X_{i+1})}{2}$;
- The point corresponding to L_i and X_i will be replaced by a new point drawing from the remaining prior volume $(0, X_i)$, whose L must be strictly larger than L_i ;
- The above procedure will be repeated until the stopping criterion is met.

To illustrate this implementation procedure, an example (also the figure for comprehension, displayed in Figure 3.13) taken from (*Skilling, 2006*) is given below, which starts with $N = 3$ points:

- Three samples are taken from the unconstrained prior, (i.e. from $(0,1)$), as shown on the lowest line in the right figure. They are indicated as point A, B, C in the left figure.
- The point A which has lowest L of three points is labelled as point X_1 , with likelihood L_1 .
- Point A is then replaced by a new point, point D, drawn from $(0, X_1)$ to assure its L is larger than L_1 .
- Now point B, C, D are three survivors uniformly distributed in the reduced range $(0, X_1)$.
- Since point D has the lowest L , it is then label as point 2 and replaced by point E.
- Suppose after 5 repetition the process could be terminated, there are five “replaced” points labeled as point 1, 2, 3, 4, 5 and three remaining survivors could be then labeled 6, 7, 8 to give all 8 points (X_1, \dots, X_8) shown on the top line in the right figure.
- The evidence and posterior could be estimated using these 8 points as described above.

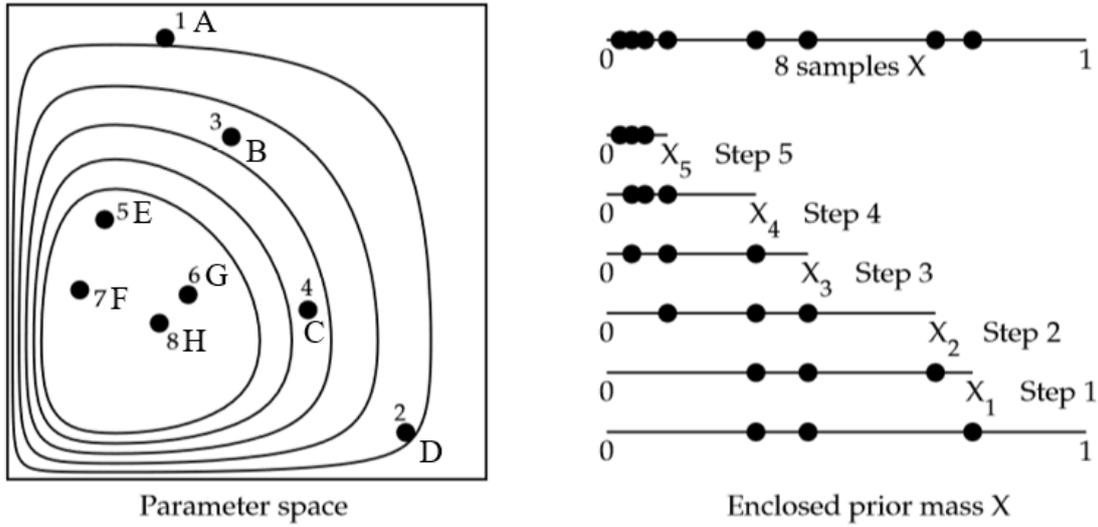


Figure 3.13: Example of nested sampling.

3.4.3.3 Stopping criterion

In this algorithm, the specific number of sampling points is not given in advance as what other MCMC methods do, a different stopping criterion is adopted. In each iteration i , the remaining evidence is estimated and once the remaining evidence over the total evidence falls below a certain threshold, the iteration would stop:

$$Z_{est} = L_{\max} X_i \quad (3.50)$$

$$\log(Z_i + Z_{est}) - \log Z_i < d \log z \quad (3.51)$$

- Z_{est} the remaining evidence at iteration i
- L_{\max} the highest likelihood in the region estimated by the current active points
- X_i the remaining prior volume.
- Z_i the currently cumulative evidence at iteration i
- $d \log z$ the threshold

3.4.3.4 From nested sampling to MultiNest

The most challenging part in implementing nested sampling is to sample uniformly from the restrained prior volume, and at the same time strictly fulfill the requirement that $L > L_i$. If the samples are drawn randomly as done in MH, the acceptance rate of new points will decrease gradually with the reduction of prior volume. To solve this problem, [Mukherjee, Parkinson,](#)

and Liddle (2006) proposed to use an ellipsoid to shrink the current prior volume and new samples will be drawn within this ellipsoid. This ellipsoid is determined by the covariance matrix of the current active points to just reach the maximum value of these points. And the ellipsoid will be expanded by an enlargement factor to avoid potential overestimation of evidence (if this is not done, the prior volume whose likelihood is below the current limit will probably not be included in the ellipsoid and thus the new sampling points will be biased to the center). Then the following procedure including the replacement of points who have the lowest likelihood is the same as original nested sampling.

However, the work up to now does not perform well in multimodal problem as argued by Shaw, Bridges, and Hobson (2007), who then improved it by separating these active points into clusters and constructing individual ellipsoid for each cluster. In this cluster-based method, there is one more potential issue that some peaks of the posterior may show a degeneracy which leads to a ‘banana’. A picture is taken from Shaw et al. (2007) in Figure 3.14 to show this case. Later Feroz et al. (2009) proposed a novel method, MultiNest to further improve their approach in (Feroz & Hobson, 2008) which could already deal with such degeneracy.

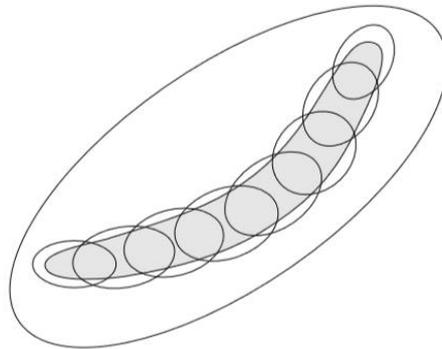


Figure 3.14: Showcase of degenerate distribution.

The essence of MultiNest is still the “cluster” based method, but now the partition of active points into clusters and construction of corresponding ellipsoid are performed simultaneously. This process is based on “expectation-minimization” scheme such that the total volumes of these ellipsoids is minimized but larger than $\frac{X_i}{e}$, where X_i is the current prior volume and

e is the desired sampling efficiency (i.e. the acceptance rate, $\frac{1}{e}$ is the enlargement factor).

After the ellipsoids are constructed, assume that at iteration i there are K ellipsoids and a particular one will be chosen from them with probability of $p_k = \frac{V_k}{V_{total}}$, where V_k is the volume

of chosen ellipsoid and $V_{total} = \sum_{k=1}^K V_k$ is the total volume of all ellipsoids. Then a sampling point could be drawn uniformly from the chosen ellipsoid. This sampling process is repeated until its L is larger than the current smallest one (i.e. L_i), where L_i is the lowest-likelihood value among all the active points at that iteration. Then this sampling point is kept with a probability of $\frac{1}{n_e}$, where n_e is number of ellipsoids in which the sampling point lies (because it is possible that these ellipsoids are overlapped). The stopping criterion could be the same as the very original nested sampling.

3.4.4 Information content

Considering the following similar questions:

- How to evaluate the performance of posterior distribution?
- How well can the data - picked up by a set of sensors - explain the state of a structure?
- How much information is carried by the data in respect of a specific state of a structure?

The term information content is introduced and three measures of it is used in this study:

- the area of the credible region;
- Shannon information entropy (hereafter just entropy for brevity);
- Kullback–Leibler divergence

3.4.4.1 Credible region

In Bayesian statistics, a credible region⁴ is a range of values within which an unobserved parameter value falls with a particular, prescribed probability. Credible regions are not unique given a specific distribution, one method is to choose the smallest region, which can be shown to coincide with an region with highest minimum probability density. Therefore, this region is sometimes called the highest posterior density region. In this thesis, always a 90% highest density credible region is used. This means that for a given distribution of a parameter, this parameter would fall into this region with a probability of 90%. Generally, the smaller the credible region, the more concentrated the posterior, which means that the parameter would approach more to a constant, and thus indicates a better parameter identification. Since most information of posterior is carried by sensors while such information can be reflected by

⁴ It is used in N-dimension case. In case of one-dimension problem, it is called credible interval.

credible region, it could be a criterion to configure sensors, including the number, the position and the type (i.e. which response should be measured).

3.4.4.2 Information entropy

The stochastic data can be considered to produce information at an average rate, such rate is called information entropy (or Shannon entropy). In case of discrete random variable θ , each possible θ value contributes to the information entropy, which is defined as:

$$H(\boldsymbol{\theta}) = -\sum_{i=1}^n p(\boldsymbol{\theta}_i) \log_n p(\boldsymbol{\theta}_i) \quad (3.52)$$

\log_n is more of a convention, n could be any number, e.g. 2, e (natural logarithm), 10, etc. In this thesis, natural logarithm will be used for convenience.

One generalization of the discrete Shannon entropy to continuous random variables is the differential entropy (or continuous entropy) which can be used as a measure of uncertainty:

$$h(\boldsymbol{\theta}) = -\int_{\boldsymbol{\theta}} p(\boldsymbol{\theta}) \cdot \log_n (p(\boldsymbol{\theta})) d\boldsymbol{\theta} \quad (3.53)$$

- $\boldsymbol{\theta}$ random variables (i.e. soil parameters to be inferred)
- $p(\boldsymbol{\theta})$ probability distribution of random variables
- $h(\boldsymbol{\theta})$ differential entropy

Papadimitriou, Beck, and Au (2000) proposed information entropy (differential entropy) to be used as the measure of information content carried by sensors. They used the measure to find optimal location of sensors. Generally speaking, the less the value of entropy, the more the information is carried.

3.4.4.3 Kullback–Leibler divergence (KL divergence)

Both credible region and Shannon entropy reflect the information conveyed by the posterior distribution, meaning that both prior and observations are included. In our work, it is preferable to quantify the information content solely from new observations (i.e. from sensors), as it explicitly provides guidance on sensor placement. Therefore, Kullback-Leibler divergence (also called relative entropy) is introduced, which is a measure of how one probability distribution is different from the other one. In our case, this KL divergence reflects additional information when moving from prior to posterior, and in contrast with Shannon entropy, the larger the KL divergence, the more the information. It is defined as:

$$H(p(\boldsymbol{\theta} | \mathbf{D}) | p(\boldsymbol{\theta})) = \int_{\boldsymbol{\theta}} p(\boldsymbol{\theta} | \mathbf{D}) \cdot \log_n \left(\frac{p(\boldsymbol{\theta} | \mathbf{D})}{p(\boldsymbol{\theta})} \right) \cdot d\boldsymbol{\theta} \quad (3.54)$$

$\boldsymbol{\theta}$	random variables (i.e. soil parameters to be inferred)
$p(\boldsymbol{\theta})$	prior distribution of random variables
$p(\boldsymbol{\theta} \mathbf{D})$	posterior distribution, an update of prior distribution
$H(p(\boldsymbol{\theta} \mathbf{D}) p(\boldsymbol{\theta}))$	KL divergence

In our work, this KL divergence reflects additional information when moving from prior to posterior, and in contrast with entropy, the larger the KL divergence, the more the information.

3.5 Reliability assessment

Structural reliability is related to the probabilistic analysis (e.g. failure probability) of engineering structures, as they usually have uncertain structural parameters (e.g. uncertain soil parameters in hydraulic structures) and uncertain behavior (e.g. nonlinear behavior of sheet pile wall). To solve this probabilistic problem, the concept of limit state was developed, by which the boundary between the safe domain and the failure domain is defined. The related function is termed as performance function $g(\boldsymbol{\theta})$.

Then the limit state is indicated by $g(\boldsymbol{\theta}) = 0$ and the failure domain is indicated by $g(\boldsymbol{\theta}) < 0$. With this definition, the failure probability can be calculated as the integral of the joint probability distribution over the failure domain:

$$P_f = P(g(\boldsymbol{\theta}) < 0) = \int_{g(\boldsymbol{\theta}) < 0} f(\boldsymbol{\theta}) \cdot d\boldsymbol{\theta} \quad (3.55)$$

Where

$f(\boldsymbol{\theta})$ joint probability distribution of $\boldsymbol{\theta}$;

P_f failure probability.

Generally, the direct calculation of the integral in equation (3.55) is computationally not tractable for high dimensional problems. Therefore, there are many methods available to approximately replace the integral, e.g. first order reliability method (FORM), second order reliability method (SORM), Monte Carlo simulation (MCS), importance sampling (IS), directional simulation (DS), etc.

However considering the posterior distribution, it is in our work in the form of representative samples without a specific distribution type, hence the mentioned methods above, e.g. FORM, IS cannot be adopted in this case. MultiNest based on nested sampling can solve this issue. When it is adopted to draw samples from the posterior density as described in section 3.4.3, each sampling point has its weight (w_i) indicated in equation (3.49). As the weight is already normalized in this algorithm, the failure probability is indeed the summation of the weights of those sampling points which fall in the failure domain:

$$P_f = \sum w_i \quad (3.56)$$

where

P_f failure probability

w_i weight of i^{th} sample in the failure domain

In our work, there might be a case where there is no clear “line” to distinguish safety and failure such that each sample in the complete space is possible to result in failure, even though of which the probability might be tiny. Therefore, the failure probability can be calculated by the following:

$$P_f = \sum p_i w_i \quad (3.57)$$

Where

P_f the failure probability;

w_i the weight of the i^{th} sample in the entire space;

p_i failure probability corresponding to each sample.

p_i is calculated considering the value or values (\mathbf{e}_i) corresponding to the sample and w_i as fixed and considering the rest of the parameters (\mathbf{R}) as random variables: $p_i = P(g(\mathbf{R}, \mathbf{E}) < 0 | \mathbf{E} = \mathbf{e}_i)$.

experience soil failure as well. The former might not be realistic while the latter is decided to be avoided in this study for simplicity. Therefore, the main properties of the three soil layer used in this study are again modified and collected in Table 4.1 (since some of these soil parameters will be assumed unknown and inferred back, the values in this table are thus named as “real” values). The soil properties are selected to avoid soil failure, in order to make further analyses easier, i.e. we always have a numerical outcome.

Table 4.1: Soil properties.

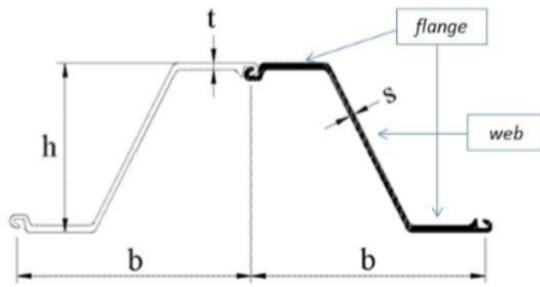
Soil type	Top soil		Mid soil		Bottom soil	
	Medium stiff clay		Medium stiff clay		Dense sand	
	Symbol	“Real” value	Symbol	“Real” value	Symbol	“Real” value
Elastic modulus [kN/m^2]	E_{top}	2850	E_{mid}	4150	E_{bot}	200000
Internal friction angle [degrees]	ϕ_{top}	19.1	ϕ_{mid}	22.7	ϕ_{bot}	45
Saturated unit weight [kN/m^3]	γ_{top}	17.8	γ_{mid}	20.3	γ_{bot}	20
Cohesion coefficient [kN/m^2]	C_{top}	18.2	C_{mid}	25.4	C_{bot}	1

The sheet pile wall is set to prevent large soil movement and the anchor reinforces the sheet pile wall. The cross-sectional properties of the sheet pile wall are given in Figure 4.2. The anchor has a diameter of 60 mm (area = 2826 mm²) with longitudinal spacing of 1.6m. The construction stages are also taken into account, which are divided into seven phases in Plaxis:

- P1: Undisturbed soil conditions with ground surface at NAP +5 m;
- P2: Excavation of top soil layer to NAP +2 m;
- P3: Placement of sheet pile wall and anchor;
- P4: Complement of the inside⁵ top soil layer to NAP +5 m;
- P5: Excavation of the outside⁶ top and mid soil layer to NAP -7.0 m;
- P6: Water level fluctuation;
- P7: Application of surface load.

⁵ Here inside refers to the part on the right hand side (behind) of sheet pile wall shown in Figure 4.1.

⁶ Here outside refers to the part on the left hand side (front) of sheet pile wall shown in Figure 4.1.



Sheet pile:			
Width Z-element	b	630	[mm]
Height	h	427	[mm]
Thickness flange	t	13	[mm]
Cross section area	A	198	[cm ² /m]
Elastic section modulus	W _{el}	2600	[cm ³ /m]

Figure 4.2: Cross-sectional properties of sheet pile [Teixeira et al. (2015)].

The element type and material model of the components, e.g. soil, anchor, sheet pile wall, and soil-structure interaction are summarized in Table 4.2, while the ground surface load intensity is 19.75 kPa as shown in Figure 4.1.

Table 4.2: Components in the FE model.

Components in FE model		Element Type	Material Model
Soil layer	top	15-node triangular element	Mohr-Coulomb soil model
	mid		
	bot		
Anchor	Head part	Mindlin beam element	Uniaxial, linear elastic
	Body part	Elastic spring element	
Sheet pile wall		Mindlin beam element	Uniaxial, linear elastic
Soil-structure interaction		Defined by pairs of nodes with identical coordinates	Based on adjacent soil

An illustrative picture of the cumulative displacements of the structure after its final construction stage is shown in Figure 4.3. The figure also shows the applied finite element mesh.

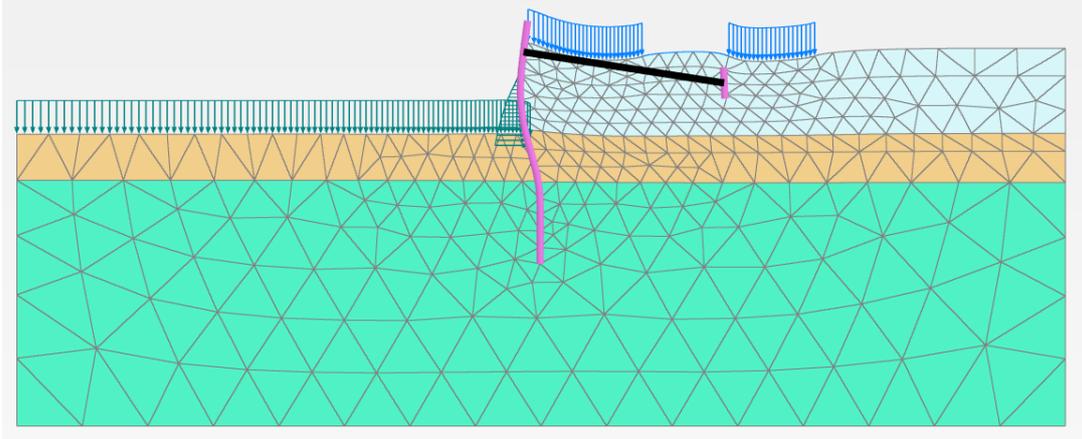


Figure 4.3: FE mesh a representative amplified, deformed shape of the analyzed sheet pile wall.

4.3 Synthetic observations

In this study, synthetic observations are used. We replace “reality” by a FE model that is also used for system identification. The usage of synthetic data has limitations, but there is also a big advantage: now we have the ground truth to evaluate the inference results. And to make the synthetic data resemble a realistic one, four kinds of observations are considered as possible measurements, namely horizontal displacement of the sheet pile wall (U_x), strain of the sheet pile wall (ε_{spw}), vertical displacement of the top ground (U_y) and strain of the anchor (ε_{anchor}).

These responses are selected as they can be easily observed and the corresponding sensors are widely used and easy to install. In practical implementation, it is not advocated to install sensors everywhere due to the limitation of budget as well as computational time. And higher sensor density does not necessarily carry more information due to the correlation among sensors in reality. Hence the sensors are selected with a space of around 2 meters in this work.

These synthetic observations generated with the Plaxis model using the “real” values in Table 4.1 are plotted in Figure 4.4, Figure 4.5 and Figure 4.6 (namely ε_{spw} , U_x and U_y), in which the cumulative responses from $P4_{start}$ to $P4_{end}$, $P5_{end}$, $P6_{end}$ and $P7_{end}$ are indicated. These generated responses are not yet contaminated with measurement uncertainty which will be included in the likelihood as described in section 3.3.3. Since the sheet pile wall is installed at $P3$, during which phase the responses related to it is tiny, it is assumed that the sheet pile wall is undeformed at $P4_{start}$. And since the top soil layer is excavated firstly and then refilled, we still assume that the top ground is undeformed at $P4_{start}$.

Therefore, we can reasonably arrange the sensors based on these figures: we do not consider to install sensors at where small magnitude of responses are. For example, strain sensors will not be installed above S#20 and below S#45 as it is expected to have small ϵ_{spw} value, and we do not install sensors below N#35 due to small U_x value.

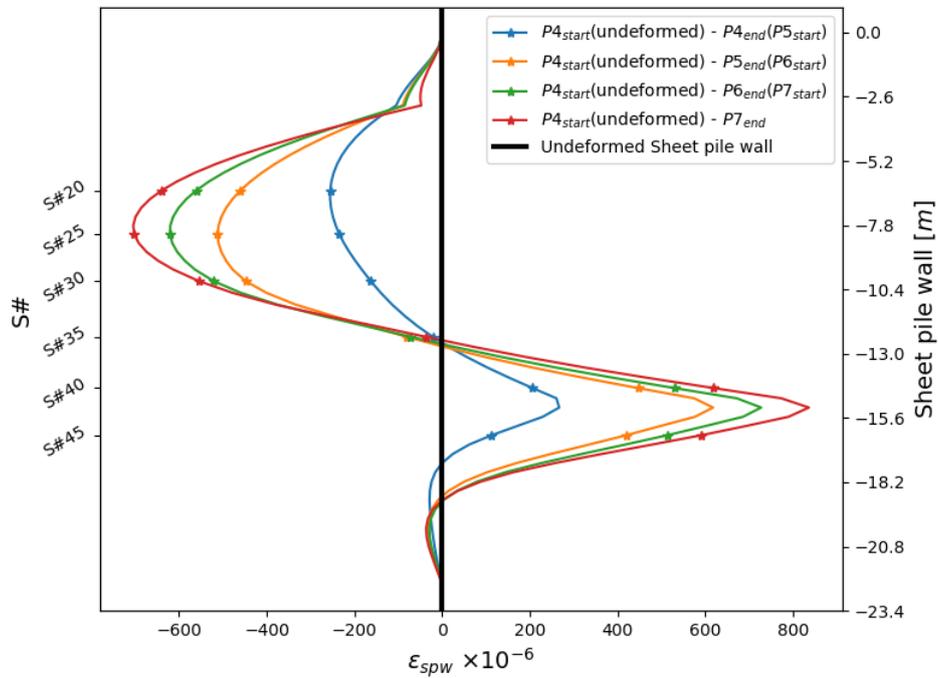


Figure 4.4: Strain of sheet pile wall corresponding to “real” values of soil properties.

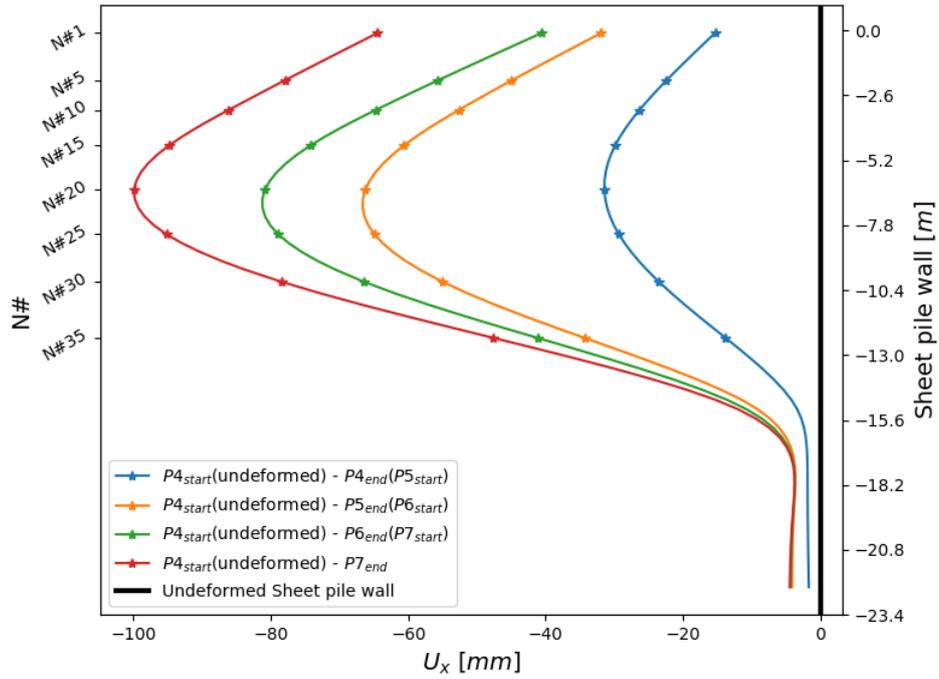


Figure 4.5: Horizontal displacement of sheet pile wall corresponding to “real” values of soil properties.

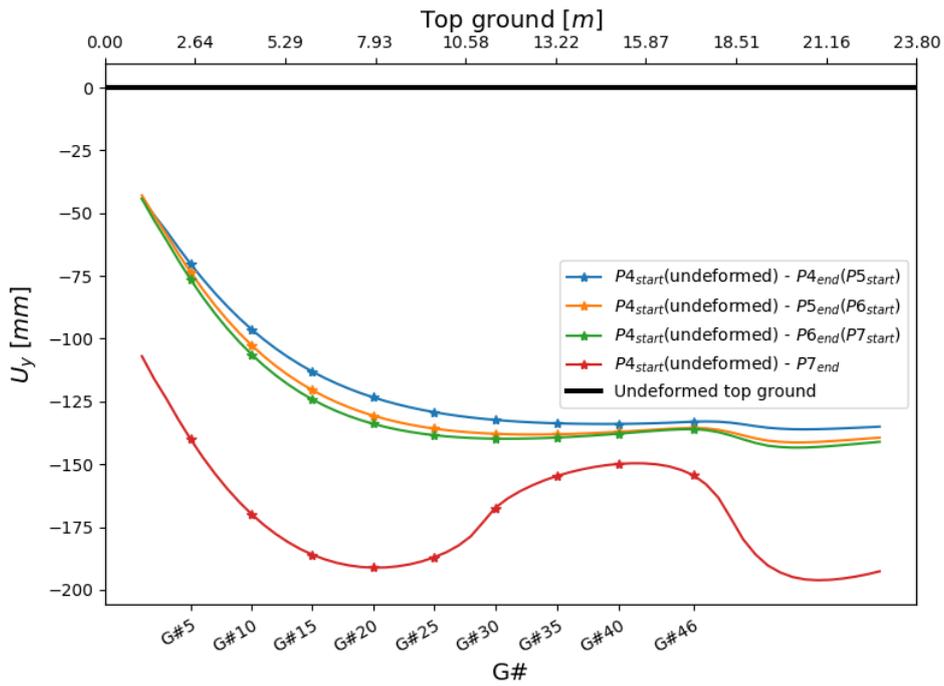


Figure 4.6: Vertical displacement of top ground corresponding to “real” values of soil properties.

Therefore, the monitoring system is indicated in a simplified drawing in Figure 4.8 where in total there are up to 24 sensors to measure different responses of this structure. Specifically,

- 8 sensors (labeled N#) measuring the horizontal displacement of the sheet pile wall (U_x) (e.g. with SAAF (SAAF, 2019));
- 6 sensors (labeled S#) measuring the strain of the outer fiber of the sheet pile wall, as shown in Figure 4.7 (ϵ_{spw}) (e.g. with strain gauge);

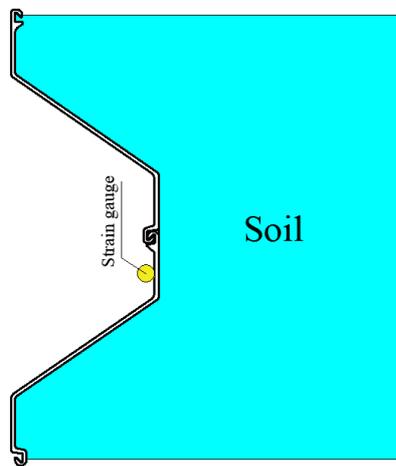


Figure 4.7: Top view of sheet pile to indicate strain gauge.

- 9 sensors (labeled G#) measuring the vertical displacement of the top ground (U_y) ;
- 1 sensors (labeled A#) measuring the strain of anchor (ϵ_{anchor}) (e.g. with strain gauge);

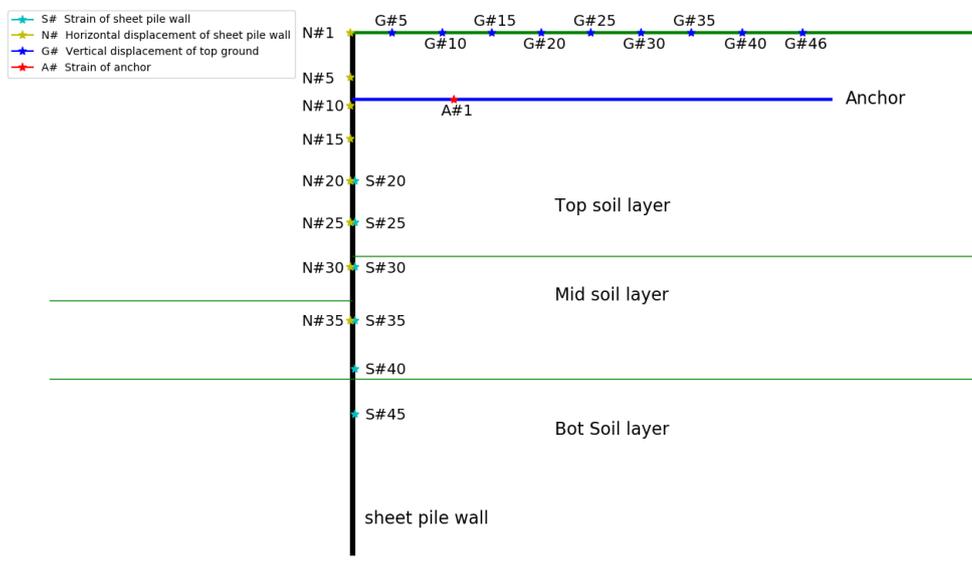


Figure 4.8: Layout of full sensor configuration.

These synthetic observations are assumed to be contaminated with noise (i.e. measurement uncertainty). The measurement uncertainty of different responses will differ and are summarized in Table 4.4.

4.4 Bayesian inference of soil parameters

4.4.1 Soil properties to be inferred

In the next step after the synthetic responses are generated, the “real” values are assumed to be unknown and the task is to infer the non-directly observable soil parameters based on the synthetic data and prior information using Bayes’ rule as shown in section 3.4.1. To perform Bayesian analysis, the properties in the top soil layer and mid soil layer (i.e. E_{top} , ϕ_{top} , γ_{top} , C_{top} , E_{mid} , ϕ_{mid} , γ_{mid} and C_{mid}) are regarded as the unknown variables, while the properties of the bottom soil layer (i.e. E_{bot} , ϕ_{bot} , γ_{bot} , and C_{bot}) are assumed to be deterministic and known.

The prior information of soil parameters is taken from (JCSS, 2001) and (Rackwitz, 2000). Since the top soil layer and mid soil layer are both cohesive soils, the same priors are assigned to them. The prior of all parameters are assumed to be lognormally distributed, as they are strictly positive. Their possible ranges of mean and coefficient of variation (COV) are given in Table 4.3. In our analysis, the means and COVs of soil parameter distributions are selected as the mean value of corresponding range in this table. The adopted range of the parameter values (i.e. support) is set between 1% cumulative probability and 99% cumulative probability (see Figure 4.9), while the tails (i.e. extreme values) are discarded to avoid soil failure in Plaxis. In this way, the prior is actually composed of truncated lognormal distributions, and the joint prior distribution is constructed by assuming mutual independence between them. Note that in Table 4.3, the selected prior mean and selected prior COV correspond to the untruncated lognormal distribution.

Table 4.3: Prior information.

	Prior mean from (JCSS, 2001)	Prior COV from (JCSS, 2001)	Selected prior mean	Selected prior COV	Support
Elastic modulus [kN/m^2] E_{top} & E_{mid}	2000 – 4000	0.2 – 1.0	3000	0.6	[708, 9350]
Internal friction angle [$degrees$] ϕ_{top} & ϕ_{mid}	19.3 – 22.8	0.1 – 0.2	21.05	0.15	[14.7, 29.5]
Saturated unit weight [kN/m^3] γ_{top} & γ_{mid}	18 – 20	0.05 – 0.1	19	0.075	[15.9, 22.6]
Cohesion coefficient [kN/m^2] C_{top} & C_{mid}	15 – 30	0.1 – 0.4	22.5	0.25	[12.3, 38.7]

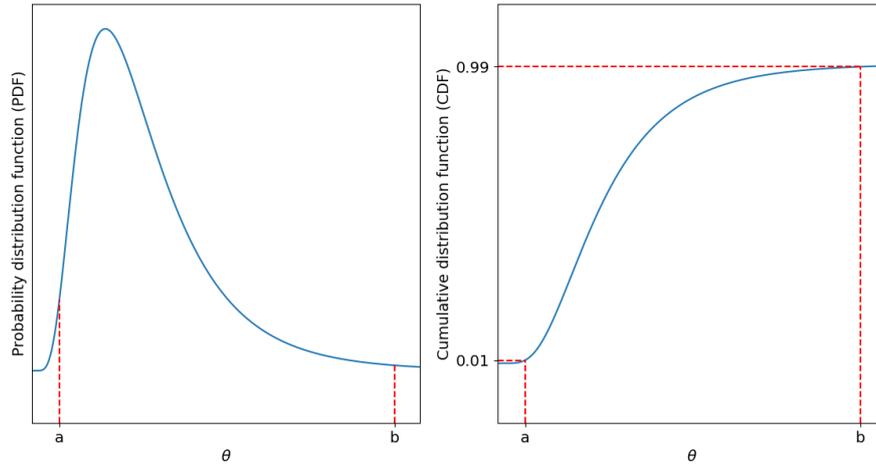


Figure 4.9: Showcase of the selection of support.

4.4.2 Uncertainties

The Bayesian inference of the soil parameters requires the evaluation of the likelihood function (see equation (3.43)), which includes three sources of uncertainties: the measurement uncertainty, the physical model uncertainty and the surrogate model uncertainty (see section 3.3). Table 4.4 summarizes the standard deviations of each uncertainty component, namely σ_{measure} , $\sigma_{\text{physical model}}$ and $\sigma_{\text{surrogate}}$ in which the σ_{measure} and $\sigma_{\text{physical model}}$ are based on expert judgement, while the $\sigma_{\text{physical model}}$ is explicitly explained by GPR. Note that due to these input values, the measurement uncertainty is expected to have a negligible effect but is still included for completeness on a conceptual level.

Table 4.4: Standard deviation of each uncertainty components.

Observation	σ_{measure}	$\sigma_{\text{physical model}}$	$\sigma_{\text{surrogate}}$
U_x [mm]	0.02	4	
$\varepsilon_{\text{spw}} [\times 10^{-6}]$	1	10	Varies with different surrogate model
U_y [mm]	0.02	4	
$\varepsilon_{\text{anchor}} [\times 10^{-6}]$	1	10	

U_x : Horizontal displacement of the sheet pile wall

ε_{spw} : Strain in the sheet pile wall

U_y : Vertical displacement of the top ground

$\varepsilon_{\text{anchor}}$: Strain of the anchor

4.4.3 Correlation in physical model uncertainty

As mentioned in section 3.2.3, the correlation in the physical model uncertainty $E_{\text{physical model}}$ is a crucial aspect in the Bayesian inference. In this numerical case study, the correlation length in $E_{\text{physical model}}$ is estimated on the basis of real measurement data from a pull-over test on a sheet pile wall and a Plaxis 3D model. The data and the Plaxis model are from (Naves, 2018). Figure 4.10 shows the finite element model, which was developed by (Naves, 2018) and has been calibrated to the experiment. Figure 4.11 shows the horizontal traction force over time that is applied at the top of the sheet pile wall. During the test, horizontal displacements of the sheet pile wall were recorded by SAAF sensors. Note that the original aim of the pull-over test is to more reliably predict the strength and deformation behavior of a combination of structural elements and (un) drained soil under extreme (high water) conditions.

The calculation of the correlation length consists the following steps:

1. The measured displacement of two SAAFs at two loading time is selected and collected;
2. The predicted displacements are generated using Plaxis positions and loading phases corresponding to SAAF measurement positions and loading phases respectively;
3. The difference between measured displacement and predicted displacement is approximated using GPR with RBF kernel along the SAAF length;
4. In total there are four length scales in four GPR models, and according to section 3.3.4, these length scales are equivalent to correlation length in physical model uncertainty. Their mean is selected as the final correlation length.

This calculation relies on the following assumptions:

- The measurement uncertainty is neglected and the difference between measured displacement and Plaxis prediction represents the physical model uncertainty;
- A GPR model is a reasonable description of observed and predicted differences;
- The measurement data at the selected two SAAFS and the time stamps are considered to be representative of the entire loading history.

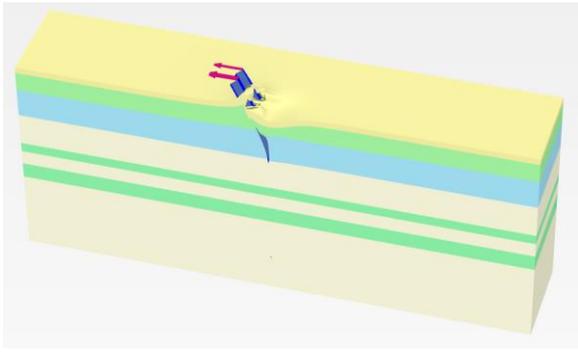


Figure 4.10: A 3D view of the FE model of the pull-over test.

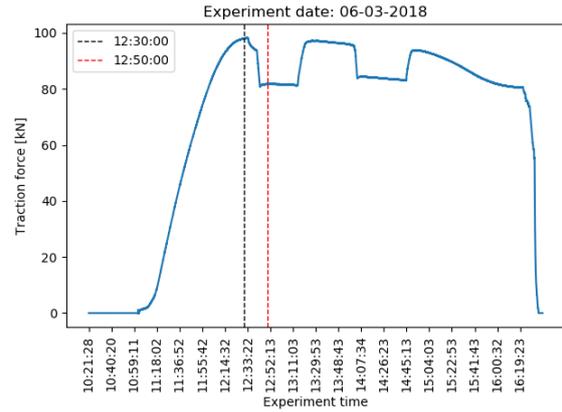


Figure 4.11: Pull-over force over time during the experiment.

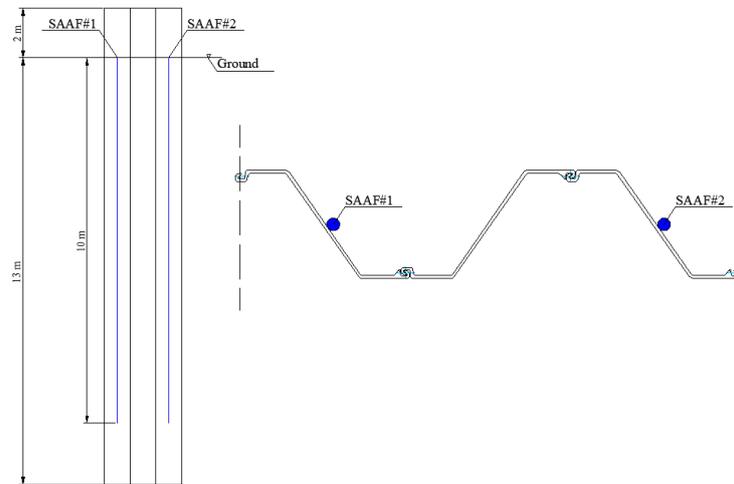


Figure 4.12: Showcase of installed SAAFs.

In our work, the horizontal displacements of two SAAFs (SAAF, 2019) are considered, which are installed along the sheet pile wall as indicated in Figure 4.12. Only the data collected at time 12:30:00 and 12:50:00 are used. The measured values and Plaxis predictions are displayed in Figure 4.13 ~ Figure 4.16. The colored areas in the figures are indicating the soil layers. Using the method described in section 3.3.4, their difference (i.e. measured-predicted, denoted as ΔU_x) is modeled as a GPR with RBF kernel. The differences along the length of the sheet pile wall and the fitted GPR models are shown in Figure 4.17 ~ Figure 4.20. Then the length scale of the RBF kernel is naturally the correlation length of physical model uncertainty. The results are collected in Table 4.5.

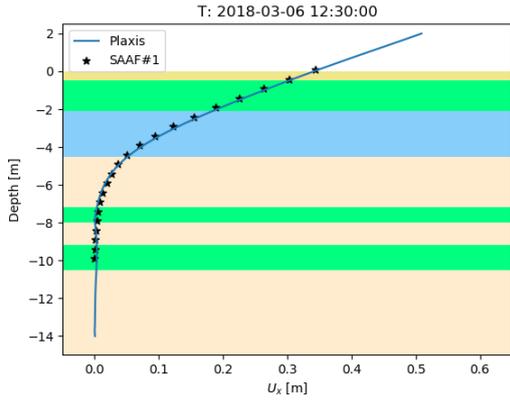


Figure 4.13: SAAF#1 measured and Plaxis predicted values at 12:30:00.

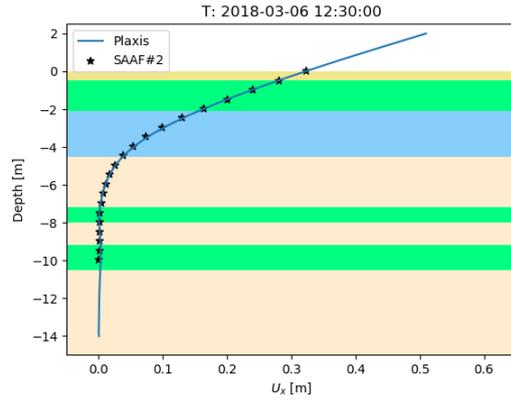


Figure 4.14: SAAF#2 measured and Plaxis predicted values at 12:30:00.

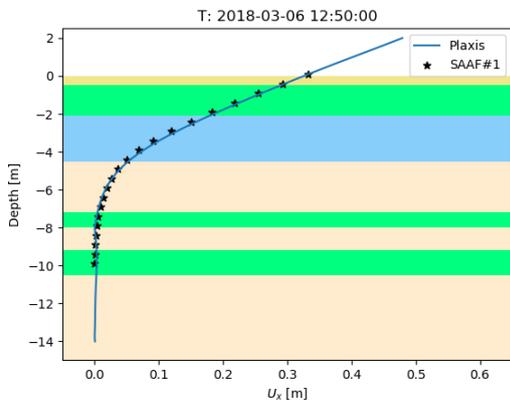


Figure 4.15: SAAF#1 measured and Plaxis predicted values at 12:50:00.

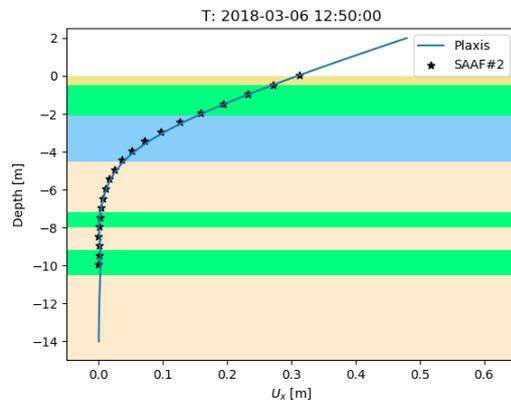


Figure 4.16: SAAF#2 measured and Plaxis predicted values at 12:50:00.

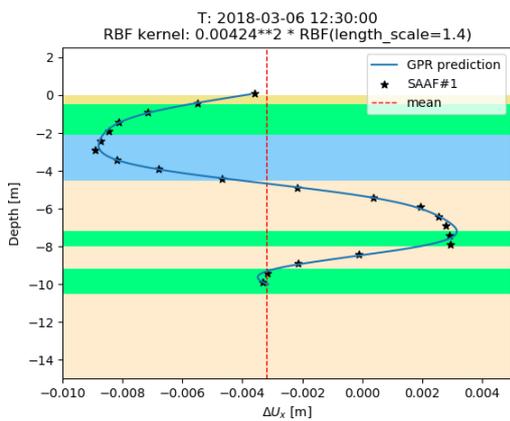


Figure 4.17: Fitting the difference between SAAF#1 measured and Plaxis predicted values at 12:30:00.

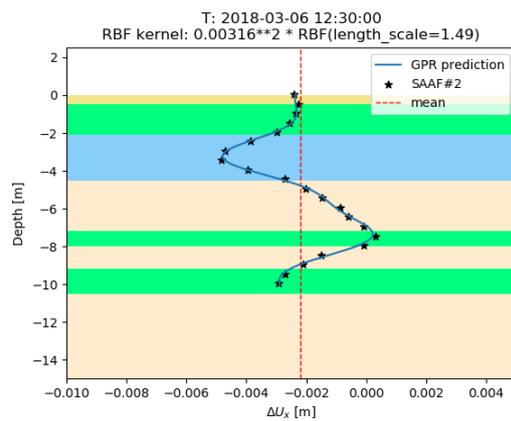


Figure 4.18: Fitting the difference between SAAF#2 measured and Plaxis predicted values at 12:30:00.

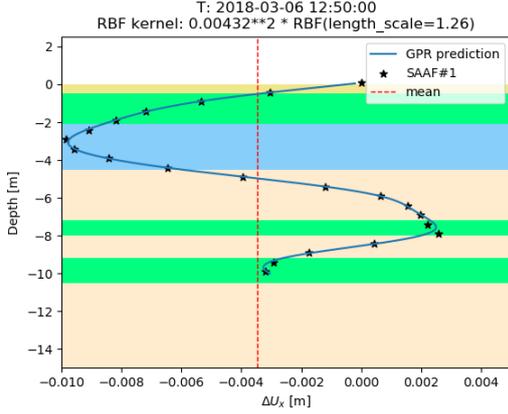


Figure 4.19: Fitting the difference between SAAF#1 measured and Plaxis predicted values at 12:50:00.

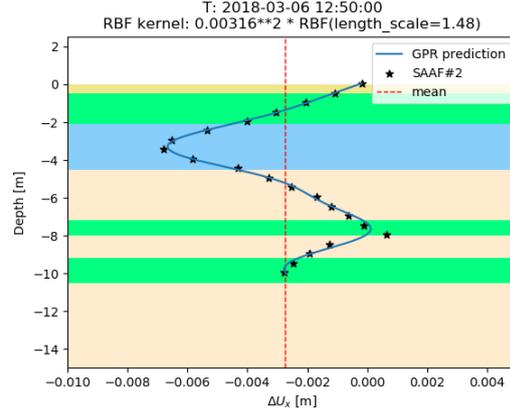


Figure 4.20: Fitting the difference between SAAF#2 measured and Plaxis predicted values at 12:50:00.

Table 4.5: Hyperparameters of the fitted GPR model.

	Mean of ΔU_x [mm]	Signal variance σ_f^2 [mm ²]	Length scale [m]
12:30:00 SAAF#1	-3.20	4.24 ²	1.40
12:30:00 SAAF#2	-2.21	3.16 ²	1.49
12:50:00 SAAF#1	-3.47	4.32 ²	1.26
12:50:00 SAAF#2	-2.74	3.16 ²	1.48
Average value	-2.91	3.76 ²	1.41

The average value (i.e. 1.41 m) is used as the correlation length of U_x in the physical model uncertainty. The use of the obtained correlation length in the 2D model entails the following assumptions:

- The physical model uncertainty of the 3D Plaxis model is the same as that of the 2D Plaxis model;
- The correlation length obtained for horizontal measurements is applicable to strain measurements and vertical measurements which are not available;
- The correlation length for physical model uncertainty also holds for surrogate model uncertainty to avoid unrealistic independence;
- The bias in the mean of ΔU_x is neglected for simplicity as it is unknown if this bias is particular for the displacement in the pull-over test.

4.5 Construction of surrogate models

To perform a single Bayesian inference, tens of thousands of runs of the Plaxis model might be needed, which is prohibitively expensive from a computational point of view. Therefore, the GPR model described in section 3.2 is adopted to construct a computationally cheap surrogate model to replace the Plaxis model. Eight soil parameters (i.e. E_{top} , ϕ_{top} , γ_{top} , C_{top} , E_{mid} , ϕ_{mid} , γ_{mid} and C_{mid}) are taken as unknown variables and hence eight dimensional surrogate models are constructed. The support of the surrogate models is the same as the support of the priors (Table 4.3). For convenience, the support is normalized to the unit hypercube in the fitting process hence all the figures presented in this report have a unit support.

Since there are 4 construction phases considered to install the monitoring system and in each there are 24 sensors, we need in total 96 surrogate models; each maps from an eight dimensional space to a one dimensional one. The details of the surrogating:

- the pseudo code of the used algorithm is described in section 3.2.6;
- stopping criterion: $\varepsilon_{\sigma} = 0.005$ is used, see section 3.2.5;
- the number of initial sampling points is set to 300 including 256 hypercube corners to speed up convergence as explained in section 3.2.7.

The first surrogate model (i.e. S#20 at P4_{start}) is constructed by running Plaxis for 628 times. Since in this process all responses generated by Plaxis are recorded, the initial sampling points will be 628 when moving to the next surrogate model (i.e. S#25 at P4_{start}). However, since some to-be-surrogated responses may have different signs (e.g. positive and negative strain of sheet pile wall) or small value of y_{range} (the range of sample values as in section 3.2.5, e.g. anchor strain at P6_{start} and P7_{start}), it can take considerable time to meet the originally set stopping criterion ($\varepsilon_{\sigma} = 0.005$). Therefore, the stopping criterion for the surrogate models at P5_{start}, P6_{start} and P7_{start} is set as $\varepsilon_{\sigma} = 0.01$. In total 1445 Plaxis runs are required to construct the 96 surrogate models with this more permissive stopping criterion for some models. The maximum of normalized prediction standard deviation excluding noise component (denoted as

$\varepsilon_{\sigma, \max} = \frac{\sigma_{\max} - noise}{y_{range}}$) for all surrogate models are shown in Figure 4.21.

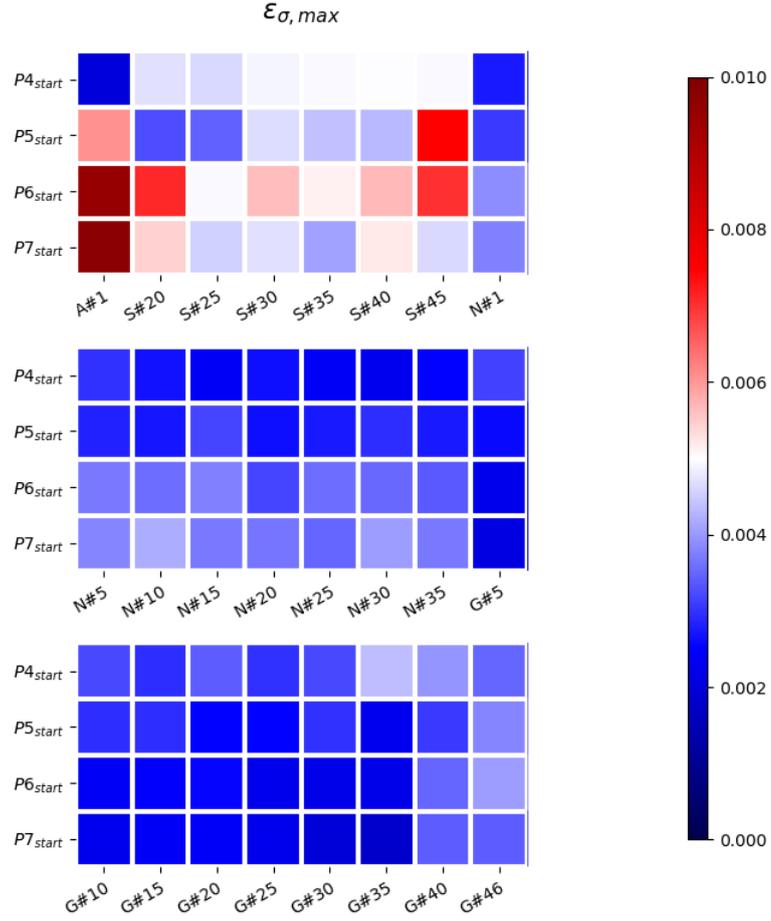


Figure 4.21: Maximum of normalized prediction standard deviation of surrogate models.

In this figure, the $\varepsilon_{\sigma, \max} < 0.005$ cases are indicated with blue squares and the darker the blue the smaller the value. While values larger than 0.005 are indicated with red squares and the darker the red the larger the value. It can be seen that the $\varepsilon_{\sigma, \max}$ of sensor N# (i.e. to measure U_x) and sensor G# (i.e. to measure U_y) are all below 0.005, while that of strain sensors (e.g. A#1 at P6_{start} and P7_{start}) are relatively higher.

To evaluate the performance of these surrogate models, 150 new sampling points (not used for the fitting) are generated with Plaxis to calculate the coefficient R^2 , which is defined as shown in equation (4.1). The results are plotted in Figure 4.22. The best score is 1.0 and the larger the score, the better model.

$$R^2 = 1 - \frac{\sum_{i=1}^{150} (y_{\text{true},i} - y_{\text{predict},i})^2}{\sum_{i=1}^{150} (y_{\text{true},i} - y_{\text{true,mean}})^2} \quad (4.1)$$

Where

$y_{\text{true},i}$ the true value generated with Plaxis of i^{th} sample;

$y_{\text{predict},i}$ the surrogate model prediction of i^{th} sample;

$y_{\text{true,mean}}$ the mean of all 150 $y_{\text{true},i}$.

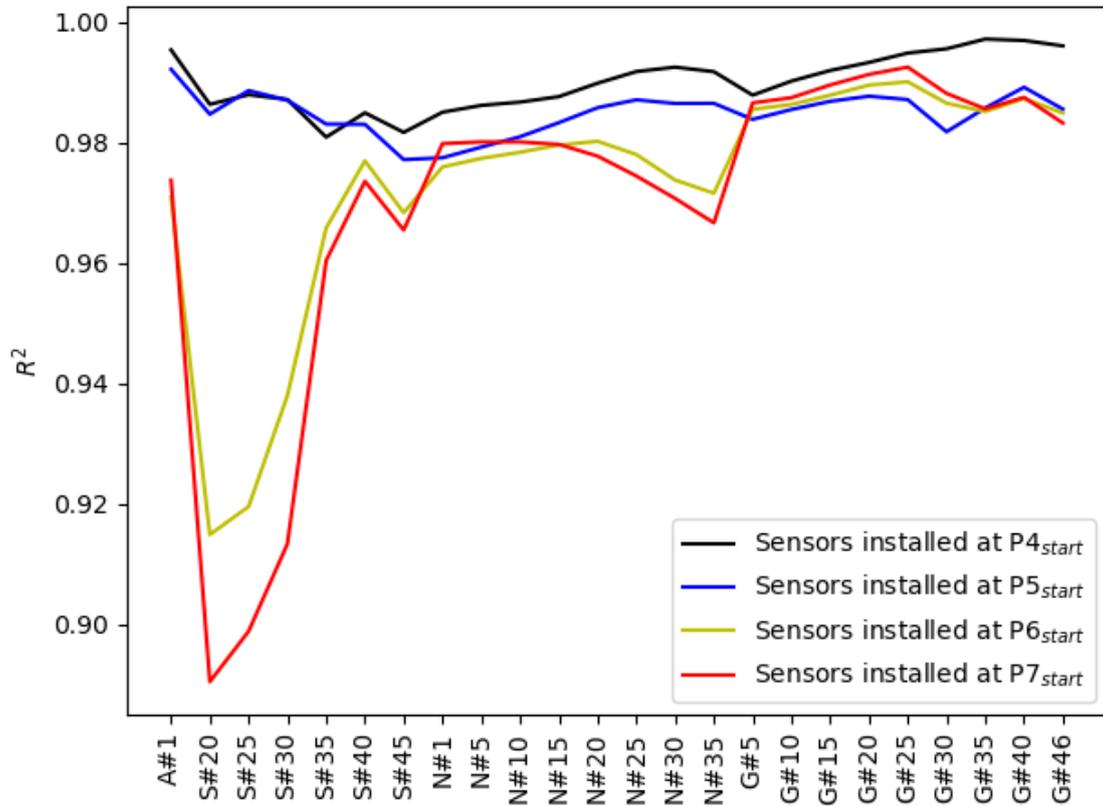


Figure 4.22: R^2 value of surrogate models.

In this figure, it is visible that the surrogate models related to S#20, S#25 and S#30 installed at P6_{start} and P7_{start} performs poorly in prediction as they have a relatively low R^2 value. This is consistent with Figure 4.21 in which these surrogate models have higher value of $\varepsilon_{\sigma,\max}$. However, a high value of $\varepsilon_{\sigma,\max}$ does not necessarily lead to a low value of R^2 as $\varepsilon_{\sigma,\max}$ is also influenced by y_{range} . For example, A#1 at P5_{start} P6_{start} and P7_{start} have very high value of $\varepsilon_{\sigma,\max}$ (i.e. close to 0.01) shown in Figure 4.21, but their prediction performance is still good reflected by the high value of R^2 .

To visualize the surrogate models and further evaluate their performance, 20 additional Plaxis runs are performed where E_{top} is selected to be the only one varied while the rest of the soil parameters are fixed at their “real” values. Then the surrogate model predictions are calculated using the same way: varied E_{top} and the rest is fixed. The responses related to S#20 installed at P7_{start} and G#35 installed at P4_{start} are given as examples as they respectively have the lowest

and the highest value of R^2 and are shown in Figure 4.23 and Figure 4.24. Note that the Plaxis outputs represented by circles in the figures were not used for fitting the GPR models.

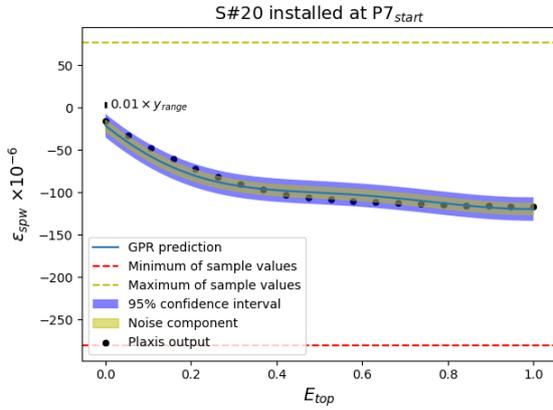


Figure 4.23: 1D visualization of S#20 at P7_start.

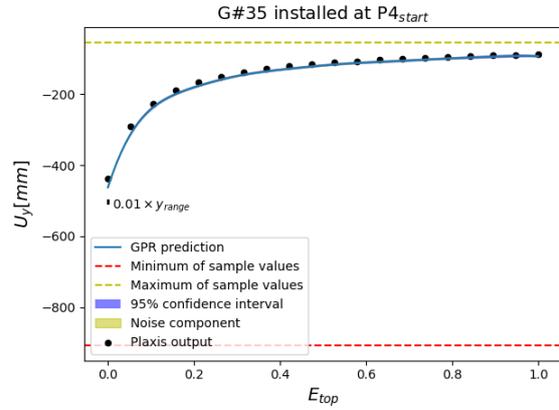


Figure 4.24: 1D visualization of G#35 at P4_start.

In both figures, the Plaxis outputs are indicated by black dots while the surrogate model prediction is indicated by the solid blue line. Additionally, the 95% confidence interval of the surrogate prediction is indicated by the blue area while the noise component is indicated by the yellow area. The range of sample values which are related to the stopping criterion is indicated by the red and yellow dashed lines and 1% of the range is indicated by the little black vertical line. In Figure 4.23 even though there is a visible difference between the Plaxis outputs and surrogate prediction, the former are all within the 95% credible region of the latter, meaning that the surrogate model is still reliable. And this figure can demonstrate that since the response may have different signs (i.e. positive or negative strain), the $\varepsilon_{\sigma, \max}$ is relatively high (corresponding to a red square in Figure 4.21) as described previously. This high $\varepsilon_{\sigma, \max}$ value thus leads to a wide 95% confidence interval. In Figure 4.24, the 95% confidence interval is quite narrow and the Plaxis output is overlapping with the surrogate prediction, which is consistent with its low value of $\varepsilon_{\sigma, \max}$ (corresponding to a blue square in Figure 4.21) and its high value of R^2 . Note that the relatively low R^2 score for some models is not an issue from a modelling point of view as the GPR model captures this discrepancy and assigns a probabilistic model to it. In later analysis, e.g. Bayesian inference, the uncertainty in the GPR model – expressed by its probabilistic model – is directly considered. The low R^2 score is more of an issue from a practical point of view as introducing additional uncertainties via the surrogate model is making our statistical inference weaker, i.e. lowers the information content we can extract from the measurements.

In the following analyses, if the number of considered unknown soil parameters is less than 8 (see Subcase 1 in section 5.1 and Subcase 2 in section 5.2), then the generated surrogate models can still be used and prediction can be made by taking the rest of the known parameters at their “real” values. Two examples are given in Figure 4.25 and Figure 4.26 considering ε_{spw} at S#20 and S#45 installed at P4_{start} and the variables are E_{top} and E_{mid} .

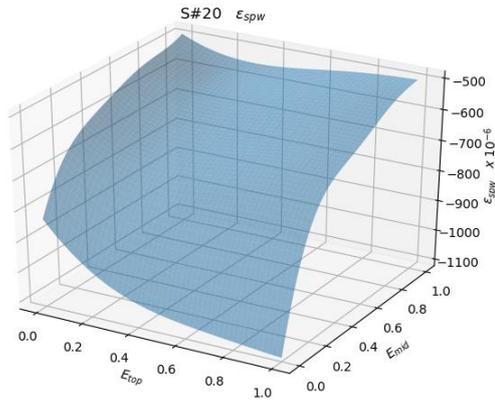


Figure 4.25: ε_{spw} at S#20 installed at P4_{start}.

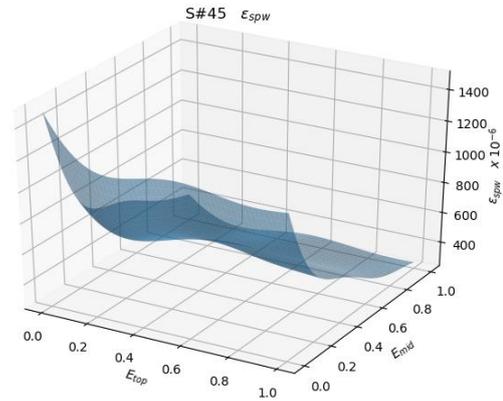


Figure 4.26: ε_{spw} at S#45 installed at P4_{start}.

4.6 Overview of analyses

The general steps of the completed analyses are the following: firstly, up to four kinds of observations are considered as possible measurements, namely horizontal displacement of the sheet pile wall, strain of the sheet pile wall, vertical displacement of the top ground, and strain of the anchor. The observations are generated with the Plaxis model. Next, surrogate models of the displacements and strains are constructed using Gaussian process regression to replace the computationally expensive Plaxis model. Then the observations (synthetic data) are used to identify soil parameters using the Bayesian approach (i.e. MultiNest algorithm). In this process, the uncertainty components, namely measurement uncertainty, physical model uncertainty and surrogate model uncertainty are considered in the likelihood function. Based on the performance of sensors, which is evaluated using KL divergence, the optimal sensor type, the optimal sensor placement, and the influence of construction stage (i.e. sensor installation time) are investigated.

Table 4.6 gives an overview of the performed analyses, namely the optimal sensor selection, the parameter identification, the influence of construction stages and the reliability analysis for three subcases with respectively two, four and eight parameters to be inferred. The next chapter presents the results of these analyses.

Table 4.6: Overview of analyses

	Case 1	Case 2	Case 3			
			Case 3-1	Case 3-2	Case 3-3	Case 3-4
Inferred parameters	E_{top}, E_{mid}	E_{top}, ϕ_{top} E_{mid}, ϕ_{mid}	$E_{top}, \phi_{top}, \gamma_{top}, C_{top}, E_{mid}, \phi_{mid}, \gamma_{mid}, C_{mid}$			
Synthetic observation	ε_{spw}	ε_{spw}	ε_{spw}	U_x	U_y	$\varepsilon_{spw}, U_x, U_y,$ ε_{anchor}
Optimal sensor selection		✓	✓	✓	✓	✓
Parameter identification	✓	✓	✓	✓	✓	✓
Influence of construction stages		✓	✓	✓	✓	✓
Reliability analysis	✓	✓				✓

5 Case study: Sheet pile wall – results

5.1 Subcase 1: E_{top} & E_{mid}

In subcase 1 the dimensionality of the inference problem is set to two. The following scenario is assumed (for an overview see Table 4.6):

- The soil parameters E_{top} and E_{mid} are considered as the unknown variables, while the other soil parameters are treated as known, deterministic values;
- The sheet pile wall has a monitoring system that comprises of six strain gauges along the height of the sheet pile (they are labelled as S#20, S#25, S#30, S#35, S#40, and S#45 in Figure 5.1);
- In this subcase, the effect of the time of installation of the monitoring system on the parameter estimation is not investigated yet, the monitoring system is assumed to be installed at the beginning of phases P4, P4_{start}⁷ (see section 4.2). The strain measurements are collected at the end of phase P7, P7_{end}⁸, in which the surface load is applied.

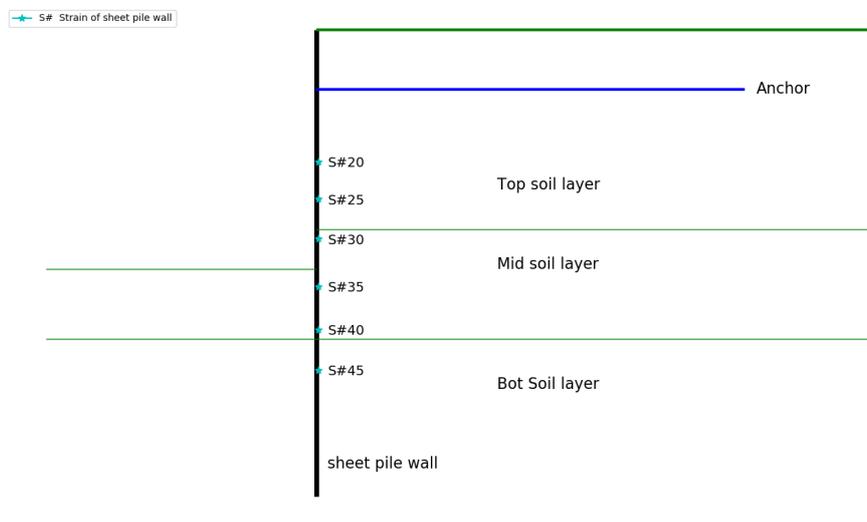


Figure 5.1: Sensor placement measuring the strain of sheet pile wall.

⁷ It means that at the start of construction phase 4, analogously, there will be P5_{start}, P6_{start}, P7_{start}.

⁸ It means that at the end of construction phase 7, analogously, there will be P5_{end}, P6_{end}, P7_{end}.

5.1.1 Parameter identification

The Bayesian computation is performed with MultiNest (see section 3.4.3), in which process the combined uncertainty is set as described in section 3.3 and the priors shown in Figure 5.2, while the posterior distribution is displayed in Figure 5.3.

In both prior and posterior plot, the red dashed lines indicate the "real" value that is aimed to be recovered from the measurements. The marginal distributions of each single parameter are plotted in the diagonal, while the joint distributions of each pair of them is plotted in the off-diagonal. The zones bounded by black lines in the off-diagonal represent 90% credible region. It can be seen that the 90% credible region of the posterior shrinks considerably compared with that of the prior, and both E_{top} and E_{mid} are well identified as their marginal distributions are quite concentrated around the "real" value in the posterior.

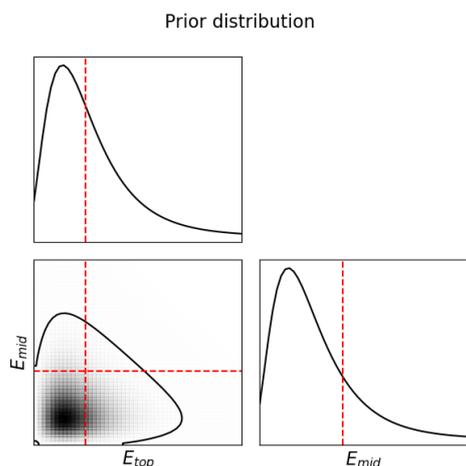


Figure 5.2: Prior distribution of Subcase 1.

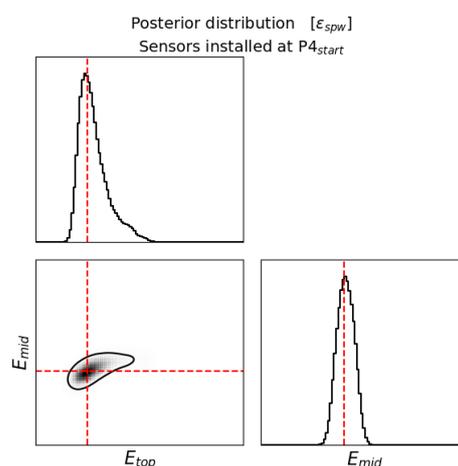


Figure 5.3: Posterior distribution of Subcase 1.

For further comparison, the mean, COV, as well as the area of 90% credible region (90% ACR) of E_{top} and E_{mid} in both prior and posterior are given Table 5.1. It is observed that that the 90% ACR of E_{top} and E_{mid} of posterior is only around 4% of that of prior. The posterior mean of both E_{top} and E_{mid} are very close to their "real" value. The COV of posterior is only 40%/15% of that of prior for E_{top} and E_{mid} respectively.

Table 5.1: Comparison of E_{top} - E_{mid} between prior and posterior.

	E_{top}			E_{mid}			90% ACR of E_{top} - E_{mid}
	"Real" value	Mean	COV	"Real" value	Mean	COV	
Prior	0.248	0.374	0.6	0.398	0.374	0.6	0.3531
Posterior		0.277	0.244		0.406	0.091	0.0149

5.1.2 Reliability assessment

In this section we investigate the effect of using updated probabilistic models of the soil parameters (posteriors) in the reliability assessment of the sheet pile wall. For this purpose, structural reliability is calculated using the prior and posterior from section 5.1.1. A serviceability limit state is considered that expresses the exceedance of a deterministic displacement threshold and thus no more random variables are required. The maximum horizontal displacement of the sheet pile wall is included in the performance function:

$$g_{SLS}(\boldsymbol{\theta}) = \hat{U}_x - \max(U_{x,i}(\boldsymbol{\theta})) \quad (5.1)$$

Where

\hat{U}_x Maximum allowed horizontal displacement before failure (deterministic);

$U_{x,i}(\boldsymbol{\theta})$ $i = 1, 5, 10, 15, 20, 25, 30, 35$, represents cumulative horizontal displacement at N#1, N#5, N#10, N#15, N#20, N#25, N#30, N#35 as shown in Figure 4.8.

Since the reliability analysis should start from the very beginning, the used responses (i.e. \hat{U}_x , $U_{x,i}(\boldsymbol{\theta})$ in this case) should be cumulative responses from when the structure is constructed to when the structure is assessed (in our case, it is from P4_{start} to P7_{end}). In this work, the \hat{U}_x is selected as 180mm, such that the failure probability for the prior distribution is around 10% (calculated using nested sampling and directional sampling). This failure probability corresponds to a reliability index of 1.3 which means relatively high cost of safety measure (see table 4.2 in (Honfi, 2013)). Based on expert judgement, we assume that the maximum of horizontal displacement of sheet pile wall occurs above N#35 and only at $U_{x,i}(\boldsymbol{\theta})$ as described. In this way, the surrogate models for Bayesian inference can be reused. Then as described in section 3.5, the failure probabilities calculated using the prior (by directional simulation) and posterior (nested sampling) are collected in Table 5.2.

From this table, it can be seen that the posterior failure probability is sharply decreased to a negligible value of 5.11×10^{-88} compared with the prior failure probability: 9.72×10^{-2} . To gain further insight into this significant reduction, the projections of the failure region for both prior and posterior are plotted in Figure 5.4 and Figure 5.5. The red zones indicate the projection of the failure region and the darker the color, the larger the exceedance of allowed horizontal displacement. The black lines and red dashed lines are the same as in Figure 5.2 and Figure 5.3. The observed significant difference in prior and posterior failure probabilities can be explained

with the help of the projections: a considerable area of the projection falls within the 90% credible region of the prior (Figure 5.4) leading to a large failure probability, while the projection is far away from the 90% credible region of the posterior (Figure 5.5) resulting a tiny failure probability.

Table 5.2: Prior and posterior failure probabilities in Subcase 1.

	Prior	Posterior
Failure probability	9.72×10^{-2}	5.11×10^{-88}

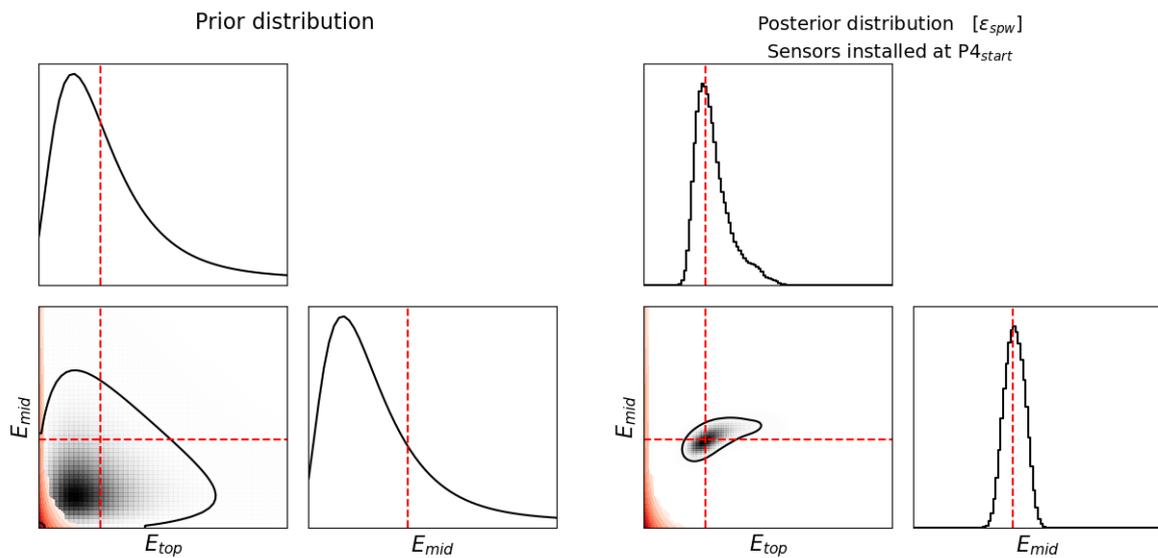


Figure 5.4: Failure region in prior distribution of Subcase 1.

Figure 5.5: Failure region in posterior distribution of Subcase 1.

5.2 Subcase 2: E_{top} , ϕ_{top} & E_{mid} , ϕ_{mid}

In Subcase 2 the dimensionality of the inference problem is set to four. The following scenario is assumed (for an overview see Table 4.6):

- The soil parameters E_{top} , E_{mid} , ϕ_{top} and ϕ_{mid} are considered as the unknown variables, while the other soil parameters are treated as known, deterministic values;
- The sheet pile wall has a monitoring system that comprises of six strain gauges along the height of the sheet pile (they are labelled as S#20, S#25, S#30, S#35, S#40, and S#45 in Figure 5.1);
- In this subcase, the effect of the time of installation of the monitoring system on the parameter estimation is investigated, the monitoring system is assumed to be installed

at, $P4_{\text{start}}$, $P5_{\text{start}}$, $P6_{\text{start}}$, $P7_{\text{start}}$ (see section 4.2). The strain measurements are collected at $P7_{\text{end}}$ in which the surface load is applied.

5.2.1 Parameter identification

Firstly, all six sensors are considered and they are assumed to be installed at $P7_{\text{start}}$. Then Bayesian analysis is conducted similarly as in section 5.1.1. The used prior distribution and obtained posterior distribution are shown in Figure 5.6 and Figure 5.7. In these two figures, the red lines indicate “real” values, and the 2D marginals and 1D marginals are plotted in the off-diagonal squares and the diagonal squares, while the black lines in 2D marginals indicate the 90% credible regions.

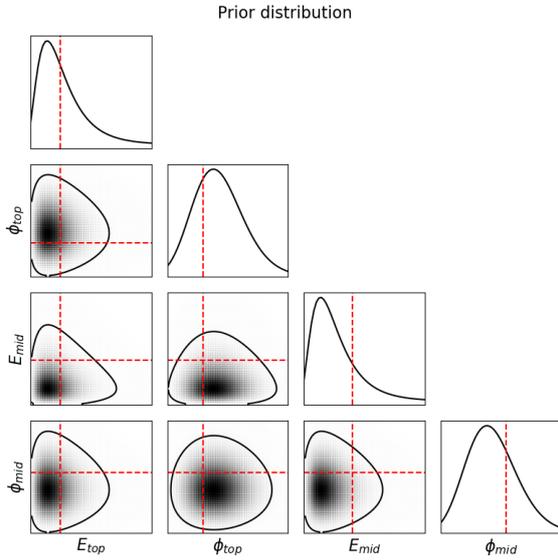


Figure 5.6: Prior distribution of Case 2.

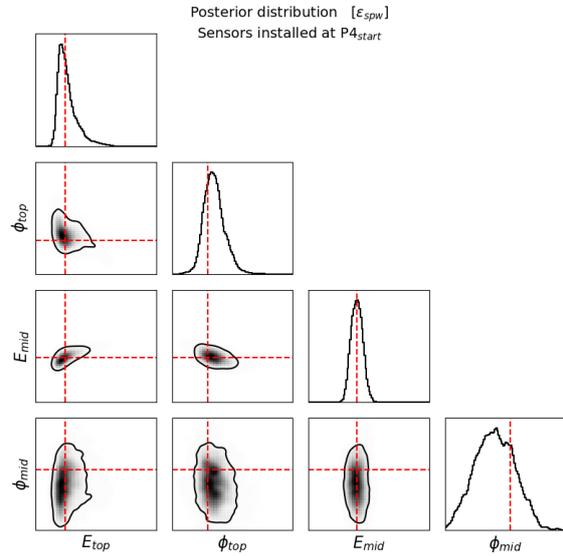


Figure 5.7: Posterior distribution of Case 2.

Based on the posterior distribution in Figure 5.7, E_{top} , ϕ_{top} and E_{mid} can be relatively accurately identified, as their marginal distributions are concentrated around a certain value. For a comparison between this subcase and Subcase 1: Table 5.1 is extended to include the results of E_{top} and E_{mid} of Subcase 2 and shown in Table 5.3. Comparing the prior and the posterior in Subcase 2, both E_{top} and E_{mid} can be properly identified (the posterior mean is close to the “real” value), and the COV of E_{top} and E_{mid} in posterior decreases to 50%/18% of that in prior. However, compared with the posterior in Subcase 1 and that in Subcase 2, there is a slight increase of the marginal COV and an almost doubled 90% ACR. This can be explained by the increasing dimensionality of the inference problem: the parameter space is larger and the

probability mass is spread among a larger number of likely parameter combinations that match with the measured responses.

Table 5.3: Comparison of $E_{top} - E_{mid}$ between prior and posterior.

	E_{top}			E_{mid}			90% ACR of $E_{top} - E_{mid}$
	“Real” value	Mean	COV	“Real” value	Mean	COV	
Prior		0.374	0.6		0.374	0.6	0.3531
Posterior (Subcase 1)	0.248	0.277	0.244	0.398	0.406	0.091	0.0149
Posterior (Subcase 2)		0.263	0.298		0.401	0.108	0.0263

Another observation is very little information is gained about ϕ_{mid} (~not identifiable): its posterior marginal distribution is very much like its prior. This is caused by the mechanical behavior of this structure, i.e. the strain of the sheet pile wall is hardly influenced by ϕ_{mid} as shown in Figure 5.8. In this figure, the responses are calculated by keeping all parameters at their “real” value except one which is varied. The “real” values are indicated by red dashed lines. Even though the influence of variation of more than one parameters is not shown, this figure is still insightful as it roughly shows if the responses are sensitive to changes in the parameters.

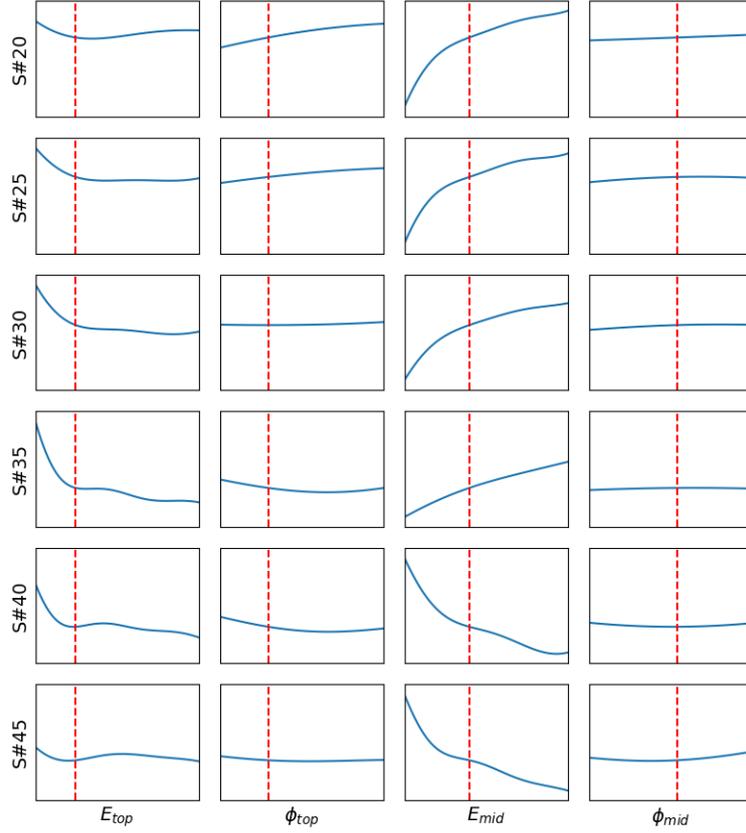


Figure 5.8 Responses of sensors installed at $P4_{start}$ in Subcase 2.

5.2.2 Optimal sensor selection

Even though more sensors convey more information, the improvement is diminishing due to the correlation among the sensors. Moreover, considering that installing more sensors costs more (e.g. cost of sensors, cost of labor considering installation and maintenance) and will largely increase computational time, it is advocated to find an optimal set of sensors and to discard less informative ones. Here the optimal set is defined as a set of sensors that conveys the most information with a fixed number of sensors. To start with, a Bayesian inference calculation is performed using different sets of observations with a combination of a single sensor and sensor pairs. For each calculation, the KLD between prior and posterior (see section 3.4.4.3) is determined to evaluate the information content of the considered sensor or sensor pair. The obtained information content is shown in Figure 5.9, in which each square of the lower triangular matrix plot corresponds to a Bayesian analysis based on the synthetic ε_{spw} from the two corresponding sensors (row and column). In the diagonal only one sensor is considered in the Bayesian analysis. The larger the KLD value (i.e. the darker the red of the square), the more informative the corresponding sensors. If posterior is the same as the prior then $KLD=0$.

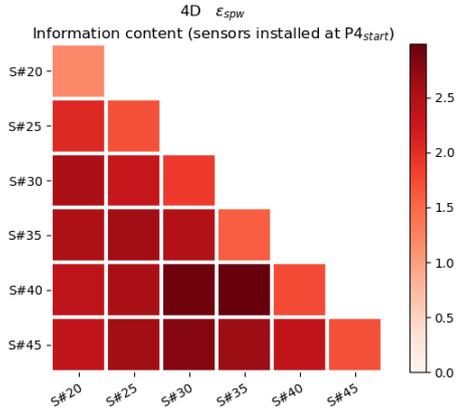


Figure 5.9: KLD matrix plot.

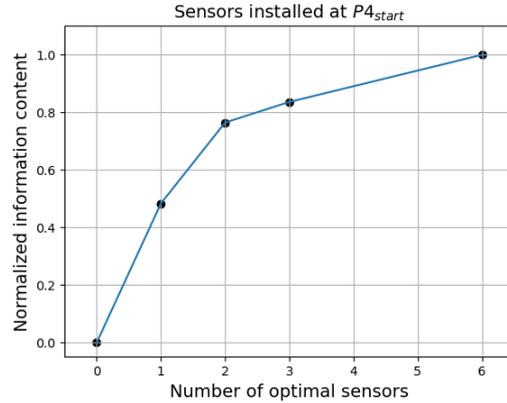


Figure 5.10: Variation of KLD with different number of sensors.

From this figure, it can be concluded that S#30 is the optimal single sensor, while S#35 and S#40 forms the optimal sensor pair. Since we are not confident if two sensors are enough, three optimal sensors is investigated following a similar analysis as for fewer sensors (i.e. the combination of six sensors taken three of them at a time without repetition). This calculation requires 20 Bayesian analyses and the obtained optimal sensors which carry the largest value of KLD are S#25, S#35 and S#40. The last Bayesian analysis uses the strains values ϵ_{spw} from all 6 sensors and quantifies the gained information using KLD. Then the normalized KLD values corresponding to the optimal sensor set for varying number of sensors are shown in Figure 5.10. It can be seen that one optimal sensor is far from sufficient to convey enough information, two optimal sensors have relatively good performance in inference, and three can convey almost 80% of the information content of six sensors. It is visible that more sensors are better, but the gained information gradually levels out with increasing number of sensors. As an example, the posterior distributions using three optimal sensors and all six sensors are displayed in Figure 5.11 and Figure 5.12 respectively. As expected, there is a small expansion of the 90% credible regions if only three sensors are used compared with usage of six sensors.

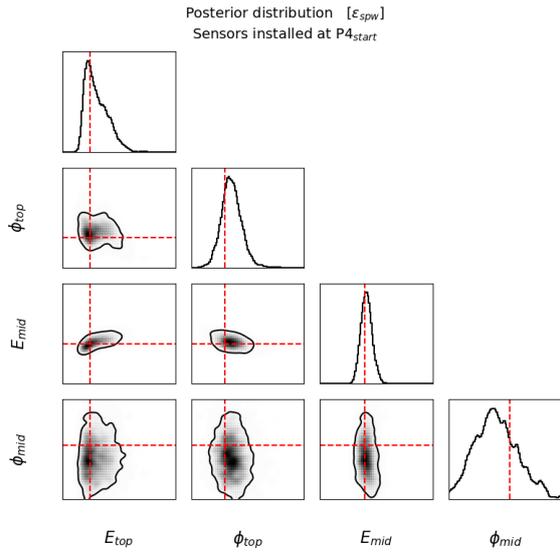


Figure 5.11: Posterior in Subcase 2 with 3 optimal sensors.

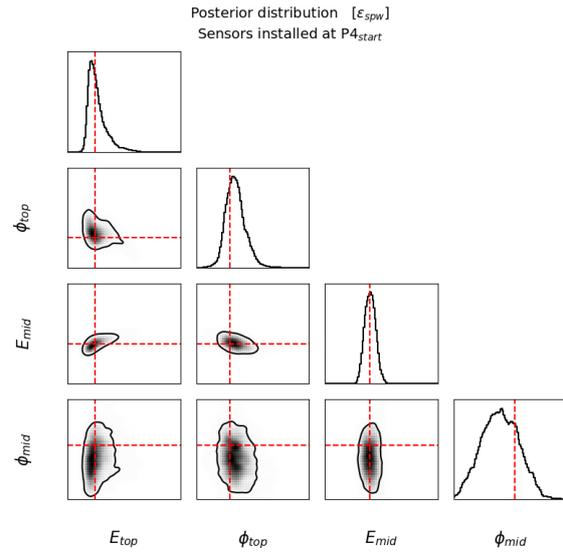


Figure 5.12: Posterior in Subcase 2 with 6 sensors.

5.2.3 Influence of construction stages

The consideration of construction phases is important to describe the soil-structure interaction in hydraulic structures accurately. In this section we analyze the effect of the starting time of sensor monitoring on the Bayesian analysis. Four different series of Bayesian calculations are performed, in which the monitoring system is installed at the beginning of phase $P4_{start}$, $P5_{start}$, $P6_{start}$ and $P7_{start}$ (see section 4.2). In all cases the strain measurements are collected at the end of phase $P7$: $P7_{end}$, when the surface load is applied. Within each series, first the three most informative strain sensors along the sheet pile wall are determined (as shown in the previous section), using their KLD matrix plots in Figure 5.13 ~ Figure 5.16. Next, the information content of the three most informative sensors per time of installation is normalized and compared in Figure 5.17.

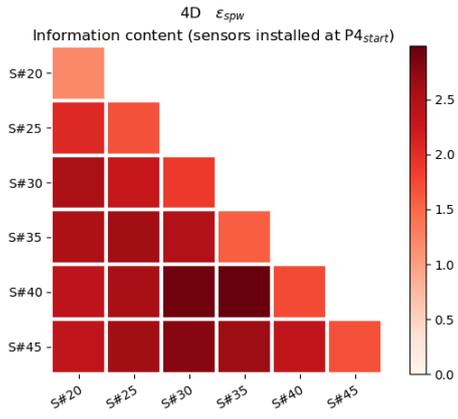


Figure 5.13: KLD matrix plot with sensors installed at $P4_{start}$,

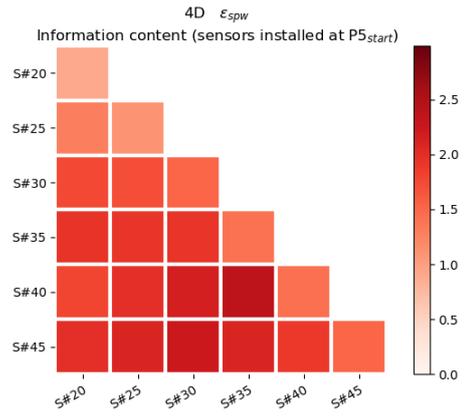


Figure 5.14: KLD matrix plot with sensors installed at $P5_{start}$,

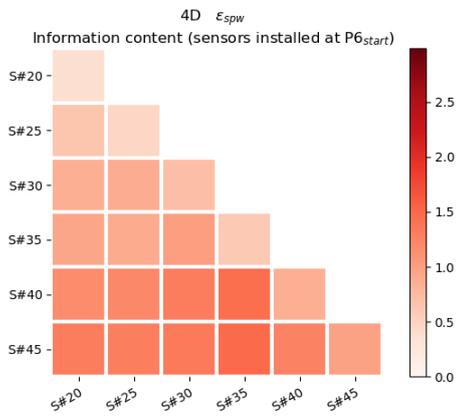


Figure 5.15: KLD matrix plot with sensors installed at $P6_{start}$,

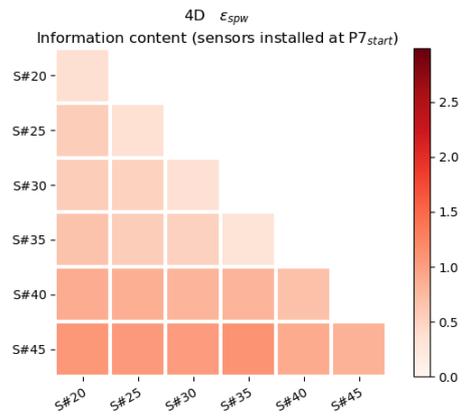


Figure 5.16: KLD matrix plot with sensors installed at $P7_{start}$,

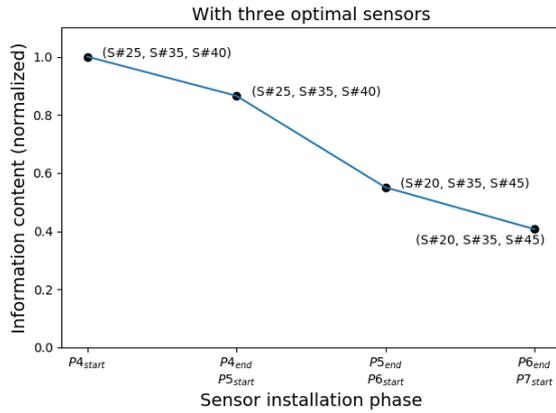


Figure 5.17: Variation of KLD with sensor installation phase. The three optimal sensors are indicated in brackets.

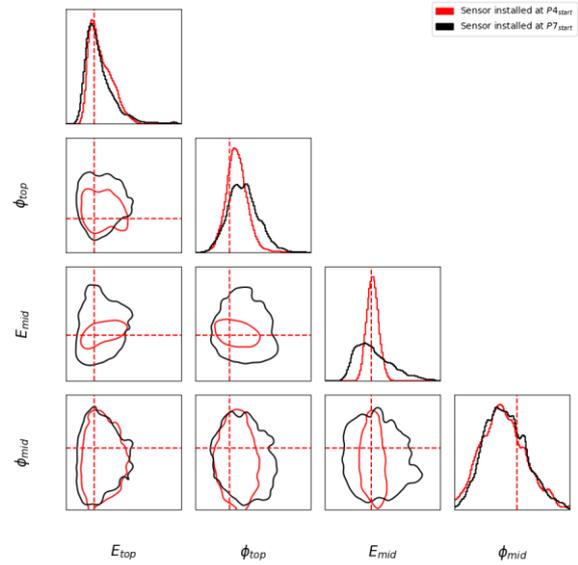


Figure 5.18: Comparison between posterior with sensors installed at $P4_{start}$ and that with sensors installed at $P7_{start}$. The contour plots in the off-diagonal plots are 90% credible regions.

From this figure, we conclude that the earlier the sensors are installed, the more information they convey regarding the estimation of the soil parameters. For example, if sensors are installed just before the surface load is applied (i.e. at $P7_{start}$), the conveyed information is only 40% of that with sensors installed at $P4_{start}$. This difference of information could also be reflected in their posterior distribution, i.e. there is a dramatic reduction of the 90% credible region with early sensor installation as shown in Figure 5.18.

5.2.4 Reliability assessment

In this section, the same serviceability limit state is applied as in section 5.1.2 (i.e. equation (5.1)). Since the aim of this section is to apply the method in section 3.5 to higher dimensional problem (i.e. 4D problem), only the prior and the posterior using sensors installed at $P4_{start}$ in section 5.2.1 are considered. Then the failure probability calculated using the prior distribution is 9.32×10^{-2} , while using the posterior it is a negligible value: 2.43×10^{-80} . Yet again we observe a sharp decrease compared with the prior value. These two failure probabilities can also be explained more straightforwardly by the projections of failure regions (i.e. the red zones) as shown in Figure 5.19 and Figure 5.20.

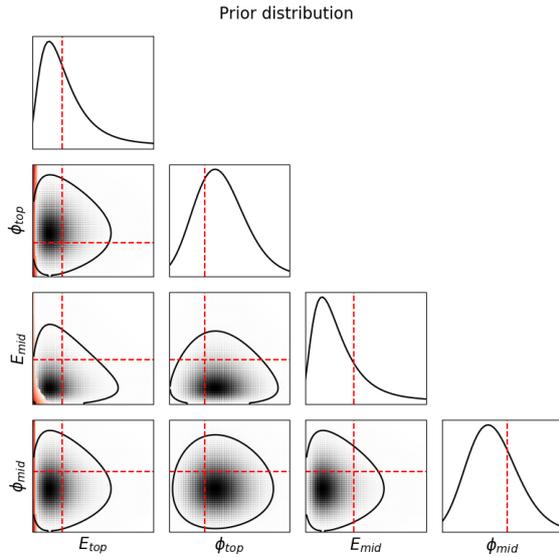


Figure 5.19: Failure projections in prior distribution.

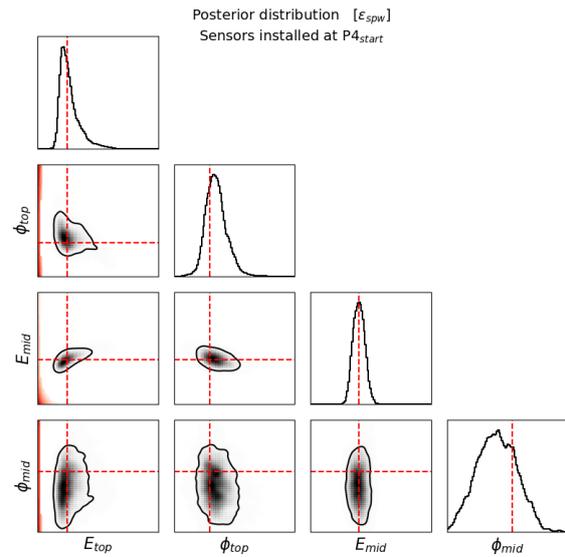


Figure 5.20: Failure projections in posterior distribution.

Since now there are 4 unknown soil parameters, the projections are made by fixing the values of non-random variables to their prior means (see section Table 4.3). For example, when the projection of failure region considering the variation of E_{top} and ϕ_{mid} (e.g. the bottom left square in Figure 5.19) is plotted, the value of E_{mid} and ϕ_{top} are fixed at the corresponding prior mean. Even though there is the limitation that the fixed parameters cannot be always at their prior mean values and the influence of variation of more than two parameters is unknown, this plotting approach is still insightful as it roughly shows how much the parameters contribute to the failure. For example, from the squares in the bottom row in Figure 5.20, it can be seen that ϕ_{mid} has little influence on the failure as the projection hardly varies with the change of it. Similarly, ϕ_{top} also has little influence and this can be confirmed from the square in the 4th row and the 2nd column where there is no failure projection, meaning that if E_{top} and E_{mid} are fixed at their prior mean, the limit state will not be exceeded regardless of the change of ϕ_{top} and ϕ_{mid} .

Based on the failure projections in Figure 5.19 and Figure 5.20, it can be explained why there is a huge decrease of failure probability: most of the failure regions are in the 90% credible region in prior, while that are considerably far away from the 90% credible region in posterior.

5.2.5 Intermediate conclusions

From the result of Subcase 2, the following intermediate conclusions are drawn:

- E_{top} , ϕ_{top} and E_{mid} can be relatively well identified using strain sensors:

- $COV_{\text{prior}}/COV_{\text{posterior}} \cong 2.5-6.5$
- The posterior mass is concentrated around the “real” values.
- ϕ_{mid} is not identifiable using strain sensors, i.e. its posterior marginal is very similar to the prior;
- With the increase of dimensionality from two to four, the COV of marginal distributions of the same parameters and the 90% ACR both increase about 20% and 75% respectively;
- It is demonstrated the KLD can be used to select the optimal set of sensors. Furthermore, it is shown that by increasing the number of strain sensors the incremental information gain is diminishing, e.g. three optimal sensors convey 80% of the information (KLD) of six sensors (the total number of sensors considered in this Subcase);
- It is advantageous to install sensors as early as possible to collect more information. The information gained when the sensors are installed before the last construction stage is only 40% of what can be obtained if they are installed at the earliest possible time;
- It has been shown that the failure probability (under serviceability limit state) can decrease by multiple order of magnitudes ($10^{-2} \rightarrow 10^{-80}$) by considering the monitoring data and using the posterior distribution compared with the prior distribution. This substantial gain is largely attributed to the substantial decrease of the scatter in the joint distribution of soil parameters. The failure probability reduction is also affected by the dimensionality of the problem: the calculated failure probability is 10^{-88} for the 2D and 10^{-80} for the 4D case after system identification.

5.3 Subcase 3: $E_{\text{top}}, \phi_{\text{top}}, Y_{\text{top}}, C_{\text{top}}$ & $E_{\text{mid}}, \phi_{\text{mid}}, Y_{\text{mid}}, C_{\text{mid}}$

In Subcase 3 the dimensionality of the inference problem is set to eight. Note that we purposefully keep the variables which have little influence on the measured responses (e.g. ϕ_{mid}) as one of the objectives of the work is to explore and test methods and tools which can be used in higher dimensional cases. Therefore the following scenario is assumed (an overview is in Table 4.6):

- The soil properties, namely $E_{\text{top}}, \phi_{\text{top}}, Y_{\text{top}}, C_{\text{top}}$ as well as $E_{\text{mid}}, \phi_{\text{mid}}, Y_{\text{mid}}, C_{\text{mid}}$ are considered as unknown variables, while the other soil parameters are treated as known, deterministic values.

- As shown in Figure 4.8, the monitoring system that we apply consists of six strain sensors on the sheet pile wall (labeled as S#), eight sensors to measure the horizontal displacements of sheet pile wall (labeled as N#), nine sensors to measure the vertical displacements of top ground (labeled as G#), and one strain sensor in the anchor (labeled as A#, since the anchor is modeled as uniaxial elastic element, the anchor force is constant along the anchor, and thus only one sensor is needed).
- The first three sets of observations are firstly considered separately in the Bayesian analysis to find an optimal set of sensors and to discard less informative ones for each of the sensor types. The performance of each observation will be compared based on KL divergence.
- The optimal sensors per sensor type are combined in one new and complete set of sensors, for which again a Bayesian analysis is performed.
- Finally, a reliability analysis is performed using the posterior distribution of soil parameters based on the complete set of optimal sensors.
- Both for the parameter estimation and reliability analysis, the calculations are performed at different times of sensor installation during construction, but all the data is collected at $P7_{\text{end}}$.
- The used prior distribution is shown in Figure 5.21, in which the black closed line indicates the 90% credible region and the red dashed lines indicate the true value to be identified. The marginal distributions of a single parameter and the joint distributions of a parameter pair are plotted in the diagonal and the off-diagonal respectively.

Prior distribution

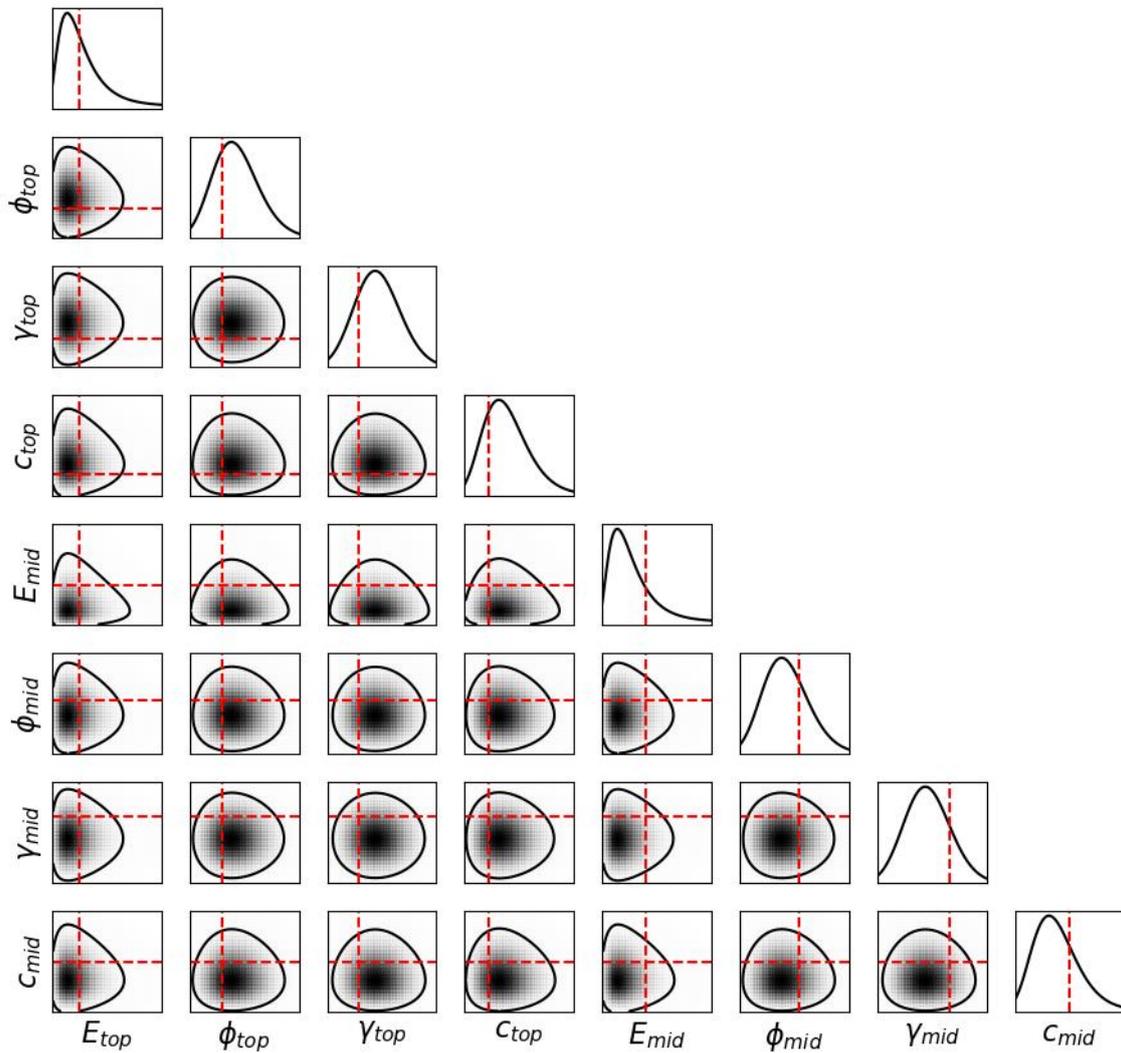


Figure 5.21: Prior in Subcase 3.

5.3.1 Optimal sensor selection

5.3.1.1 Case 3-1: Strain of sheet pile wall as observation

Firstly, only the strains of the sheet pile wall (ε_{spw}) are considered. The possible sensor locations are indicated in Figure 5.22, which is the same as what has been used in the previous Subcase 1 and Subcase 2.

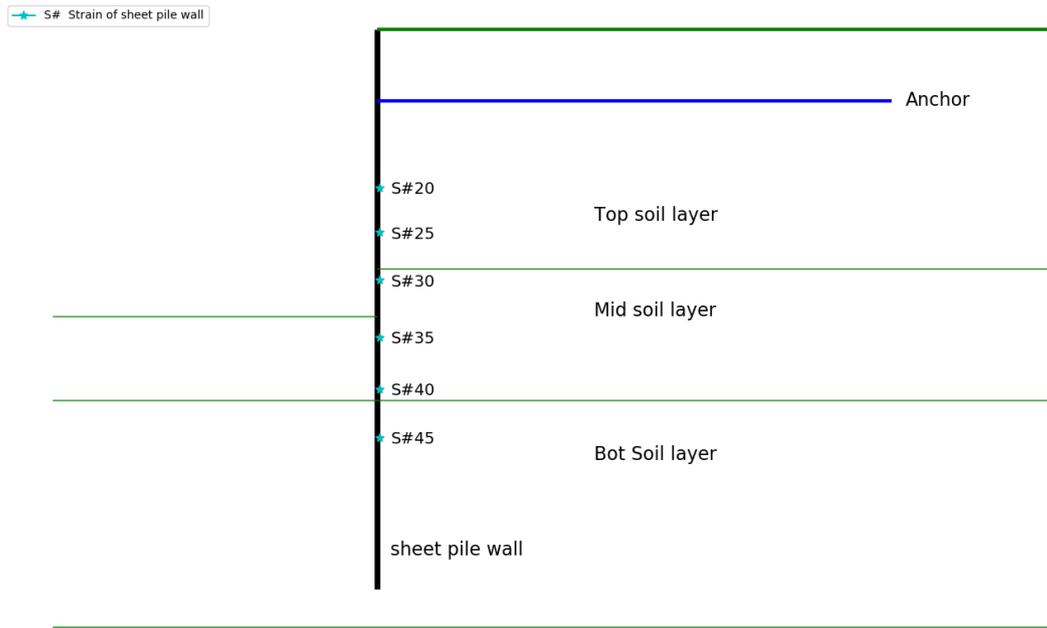


Figure 5.22: Sensor placement in Subcase 3-1.

Using the same method as in Subcase 2, the optimal sensor sets for different number of sensors are determined based on KLD values. The results are collected in Table 5.4 and the KLD values for one and two optimal sensors are displayed in Figure 5.23 ~ Figure 5.26.

Table 5.4: KLD in Case 3-1.

Sensor installation time		P4 _{start}	P5 _{start}	P6 _{start}	P7 _{start}
1 optimal sensor	Location	N#45	N#45	N#45	N#45
	KLD	2.38	2.19	1.28	0.88
	KLD (Normalized)	0.497	0.457	0.267	0.184
2 optimal sensors	Location	N#30, N#40	N#35, N#45	N#35, N#45	N#35, N#45
	KLD	3.53	3.13	2.19	1.58
	KLD (Normalized)	0.737	0.653	0.457	0.330
3 optimal sensors	Location	N#25, N#35, N#45	N#25, N#35, N#45	N#20, N#35, N#45	N#20, N#35, N#45
	KLD	4.06	3.91	2.50	1.85
	KLD (Normalized)	0.848	0.816	0.522	0.386
All sensors	Location	N#20, N#25, N#30, N#35, N#40, N#45			
	KLD	4.79	4.30	3.05	2.29
	KLD (Normalized)	1.000	0.898	0.637	0.478

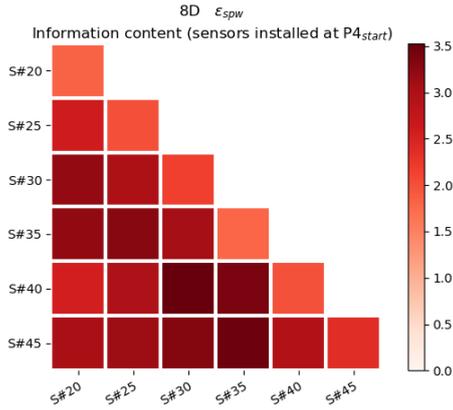


Figure 5.23: KLD matrix plot with sensors installed at $P4_{start}$.

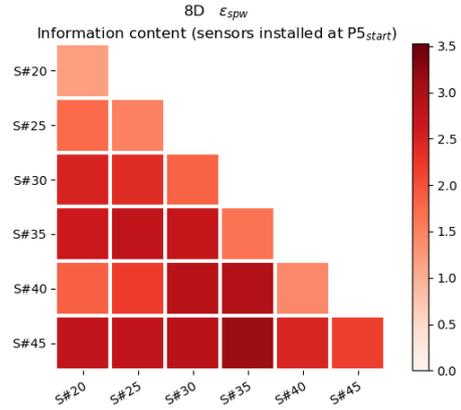


Figure 5.24: KLD matrix plot with sensors installed at $P5_{start}$.

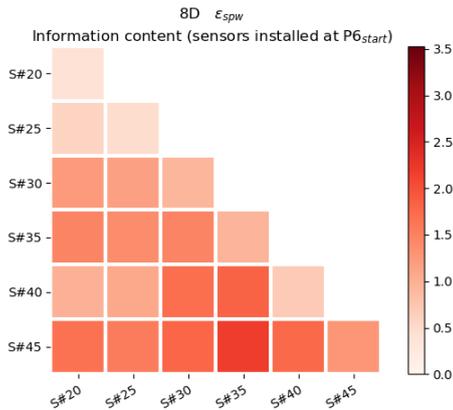


Figure 5.25: KLD matrix plot with sensors installed at $P6_{start}$.

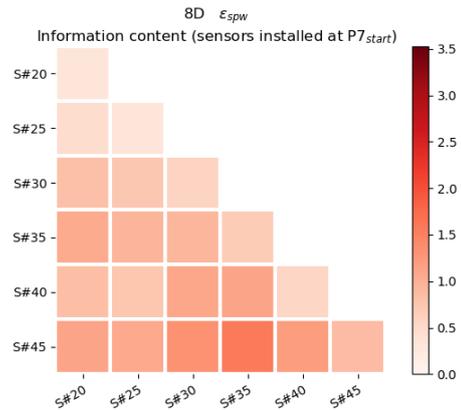


Figure 5.26: KLD matrix plot with sensors installed at $P7_{start}$.

The results are also visualized in Figure 5.27. For illustration, Figure 5.28 shows the posterior distributions of the soil parameters based on the three optimal sensors installed at $P4_{start}$. The following observations can be made from all these results:

- Three optimal sensors installed in all phases could convey sufficient information (i.e. approximately 90% of information conveyed by all six sensors installed in corresponding phase) as there is tiny correlation among the optimal sensors considering they are far away from each other;
- The optimal sensors are generally located at the height of the mid and bottom soil layer. It is in line with what we expect from a mechanical point of view, as well as from Figure 4.4 since large strain is expected at these positions;
- With each delay of sensor installation phase, there will be a considerable decrease of gained information. For example, when using all the six strain sensors, the decrease of gained information is the largest when the sensors are installed at $P6_{start}$ instead of $P5_{start}$ (around 25%). This is consistent with Figure 4.4, in which the strain of the sheet pile

wall increases the most during P5 as there is an excavation of top and mid soil layer described in section 4.2. If sensors are installed at P6_{start}, this portion of strain will not be measured and thus leading to a large decrease of KLD value;

- Even three optimal sensors installed at P4_{start} (hereafter named as optimal sensor placement) could not properly identify these soil parameters as shown in Figure 5.28. In the 2D and 4D problems (i.e. namely Subcase 1 and Subcase 2), most of the soil properties can be properly identified based on these strain measurements in the sheet pile wall. However in the 8D problem, only E_{top} and E_{mid} reveal concentrated posteriors (but their most likely value is offset a bit from the real value), the other posteriors are not so informative. A possible explanation is that for an increasing dimensionality in the inference problem, the number of parameter value combinations that match with the measured responses could increase as well. Hence adding more sensors that are not or weakly correlated to the current set of sensors could help and lead to more concentrated posteriors, which will be illustrated in case 3-4.

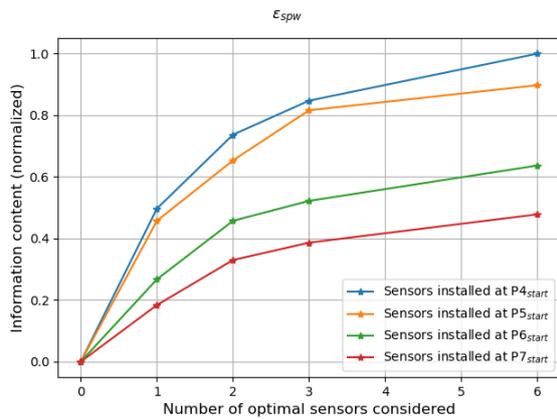


Figure 5.27: Variation of KLD in Case 3-1.

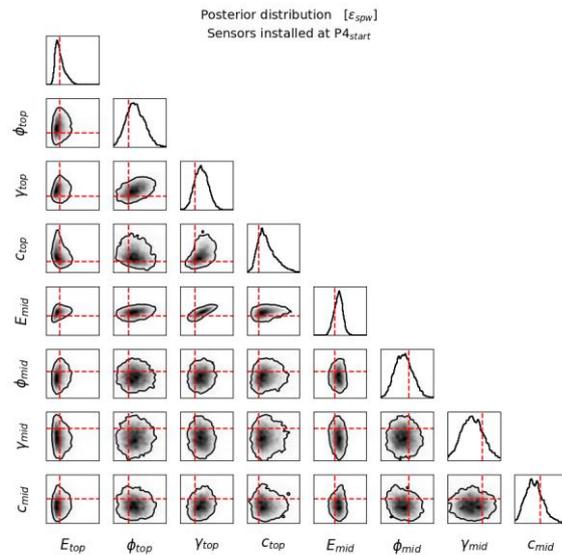


Figure 5.28: The most informative posterior in Case 3-1.

To gain more insight, the six strain sensors installed at P4_{start} are used to perform a Bayesian inference to compare the results of E_{top} and E_{mid} with that in Table 5.3. The posterior is given in Figure 5.29 and quantitative values are given in Table 5.5. From the comparison among the posterior in Subcase 1, in Subcase 2 and in Case 3-1, the COV of both E_{top} and E_{mid} as well as the 90% ACR will increase with the increase of dimensionality. This again demonstrates that

more parameter combinations are possible to match with the measured responses with an increase of dimensionality in the inference problem.

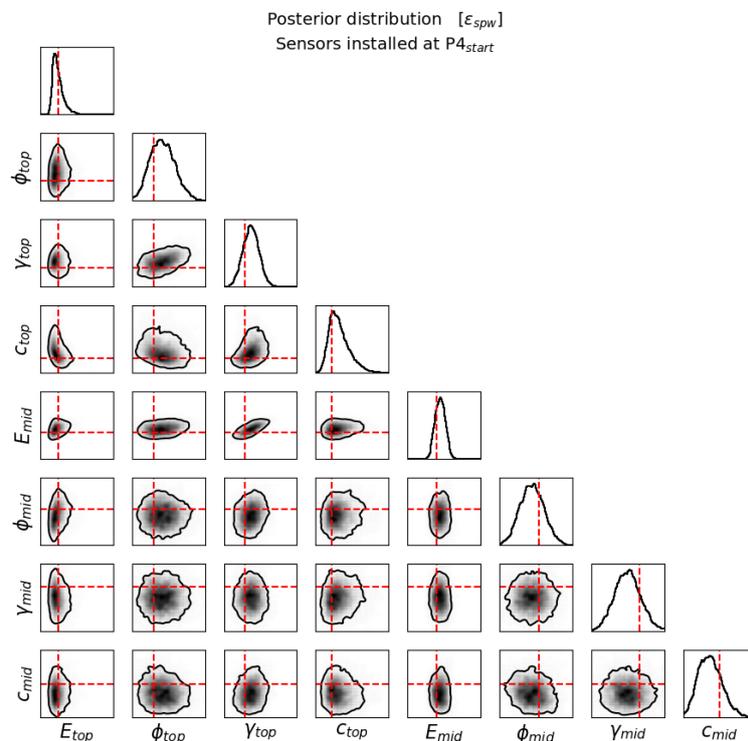


Figure 5.29: Posterior based on 6 strain sensors on sheet pile wall installed at $P4_{start}$.

Table 5.5: Comparison of $E_{top} - E_{mid}$ between prior and posterior.

	E_{top}			E_{mid}			90% ACR of $E_{top} - E_{mid}$
	“Real” value	Mean	COV	“Real” value	Mean	COV	
Prior		0.374	0.6		0.374	0.6	0.3531
Posterior (Subcase 1)	0.248	0.277	0.244	0.398	0.406	0.091	0.0149
Posterior (Subcase 2)		0.263	0.298		0.401	0.108	0.0263
Posterior (Case 3-1)		0.245	0.304		0.449	0.150	0.0688

5.3.1.2 Case 3-2: Horizontal displacement of sheet pile wall as observation

In this section, the horizontal displacements of the sheet pile wall (U_x) are used as observations to identify the soil parameters. The potential sensor configuration is shown in Figure 5.30, and the analysis performed in previous section is repeated to find the optimal sensor locations. The results are given in Table 5.7 and the KLD values for one and two optimal sensors are displayed in Figure 5.31 ~ Figure 5.34.

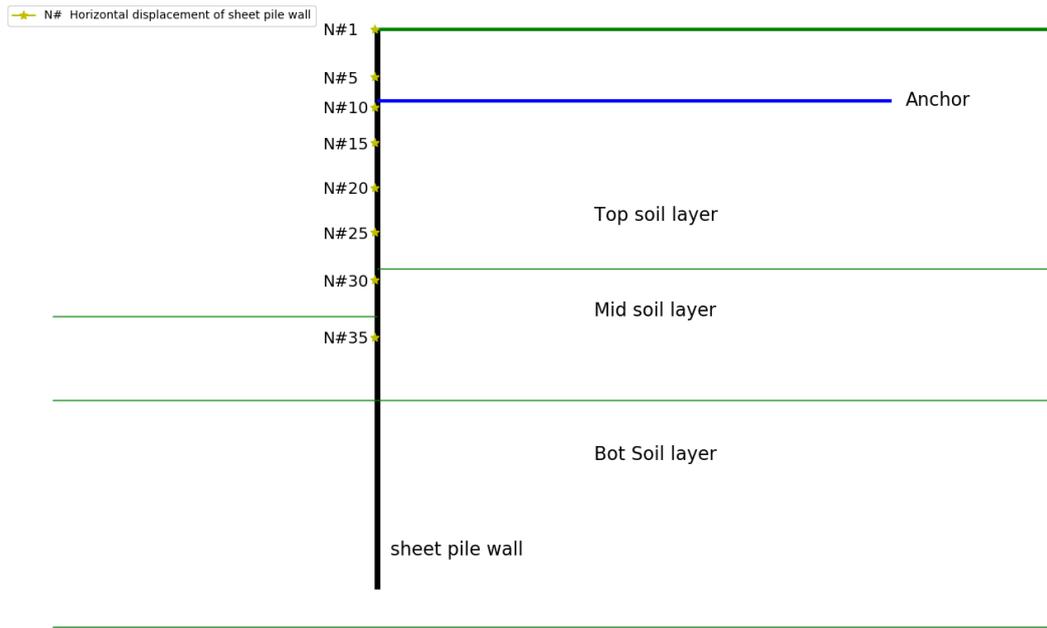


Figure 5.30: Sensor placement in Case 3-2.

Table 5.6: KLD in Case 3-2.

Sensor installation time		P4 _{start}	P5 _{start}	P6 _{start}	P7 _{start}
1 optimal sensor	Location	N#25	N#25	N#1	N#1
	KLD	1.62	1.29	0.75	0.63
	KLD (Normalized)	0.579	0.461	0.268	0.225
2 optimal sensor	Location	N#1, N#30	N#10, N#30	N#1, N#15	N#1, N#5
	KLD	2.24	1.56	1.02	0.76
	KLD (Normalized)	0.800	0.557	0.364	0.271
3 optimal sensor	Location	N#1, N#20, N#30	N#1, N#20, N#30	N#1, N#10, N#30	N#1, N#10, N#25
	KLD	2.46	1.74	1.10	0.91
	KLD (Normalized)	0.879	0.621	0.383	0.325
All sensors	Location	N#1, N#5, N#10, N#15, N#20, N#25, N#30, N#35			
	KLD	2.80	2.01	1.38	0.95
	KLD (Normalized)	1.000	0.718	0.493	0.339

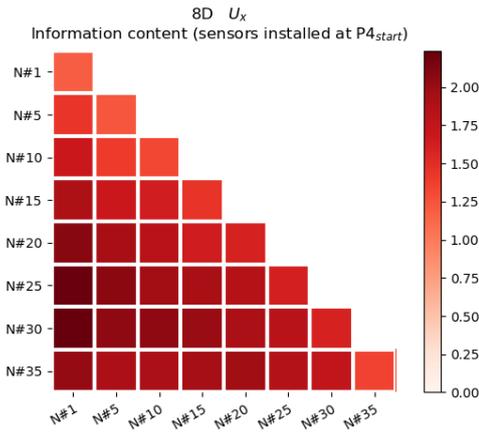


Figure 5.31 KLD matrix plot with sensors installed at $P4_{start}$.

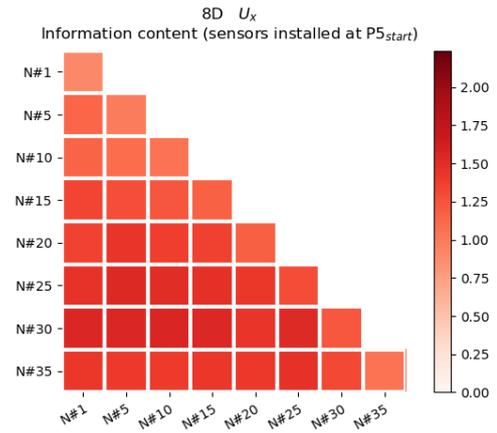


Figure 5.32 KLD matrix plot with sensors installed at $P5_{start}$.

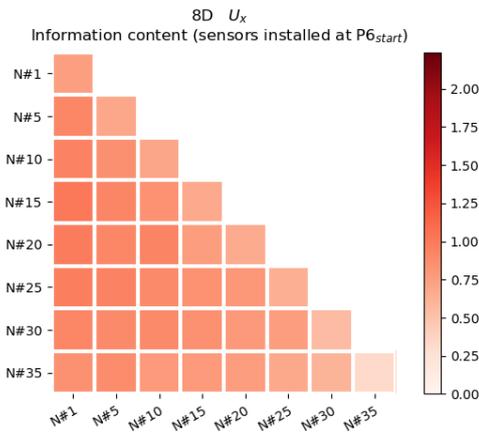


Figure 5.33 KLD matrix plot with sensors installed at $P6_{start}$.

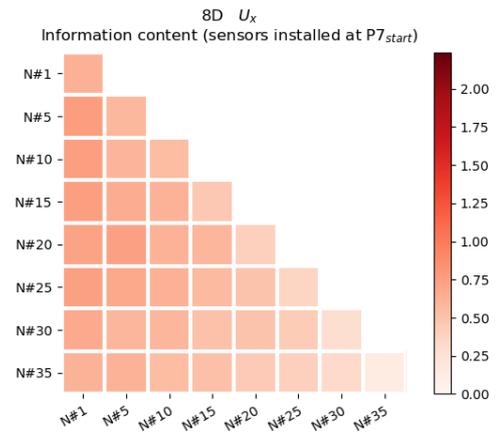


Figure 5.34: KLD matrix plot with sensors installed at $P7_{start}$.

From the summary plot in Figure 5.35, we observe similar trends as in the previous section for the strain sensors in the sheet pile wall:

- Three optimal sensors are sufficient to convey information regardless of the installation time (around 90% of information conveyed by all 8 sensors) as correlation among them is now negligible;
- The earlier the sensor installation time, the more the carried information by sensors, and every time the installation is done one stage earlier, there is a visible increase of the conveyed information (averagely 30% from $P4_{start}$ to $P5_{start}$ and from $P5_{start}$ to $P6_{start}$, even though only around 10% from $P6_{start}$ to $P7_{start}$). This quantified trend is in line with Figure 4.5, in which for example, there is considerable horizontal displacement in the $P4$ and if sensors are installed at $P5_{start}$, this portion of displacement will not be measured and leading to the large decrease of KLD value;

- Only E_{top} can be properly identified even with optimal sensor placement (i.e. three optimal sensors installed at $P4_{start}$) as shown in Figure 5.36. Considering the same reasons as discussed in Case 3-1, we hope to improve this situation by adding more uncorrelated/weekly correlated sensors.

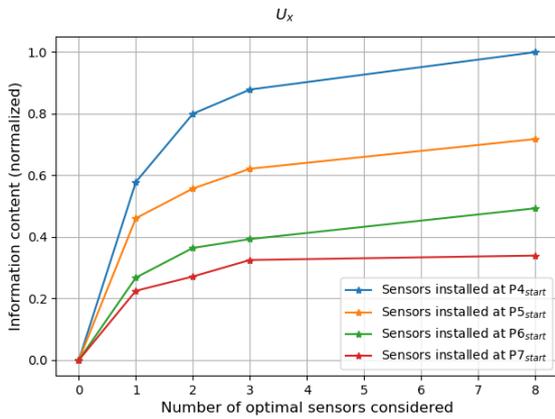


Figure 5.35: Variation of KLD in Case 3-2.

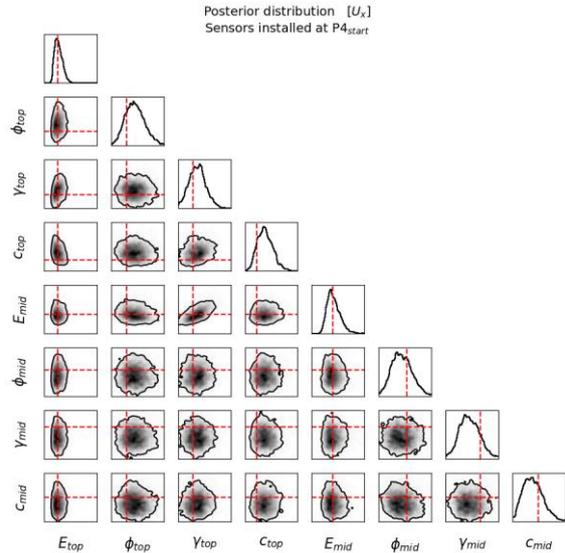


Figure 5.36: The most informative posterior in Case 3-2.

5.3.1.3 Case 3-3: Vertical displacement of top ground as observation

In this section, the vertical displacements of top ground (U_y) are considered as observations, and the sensor configuration is shown in Figure 5.37. The same analysis is performed as in Case 3-1 and Case 3-2 is repeated to find the optimal sensor locations. The results are collected in Table 5.7 and the KLD values for one and two optimal sensors are displayed in Figure 5.38 ~ Figure 5.41.

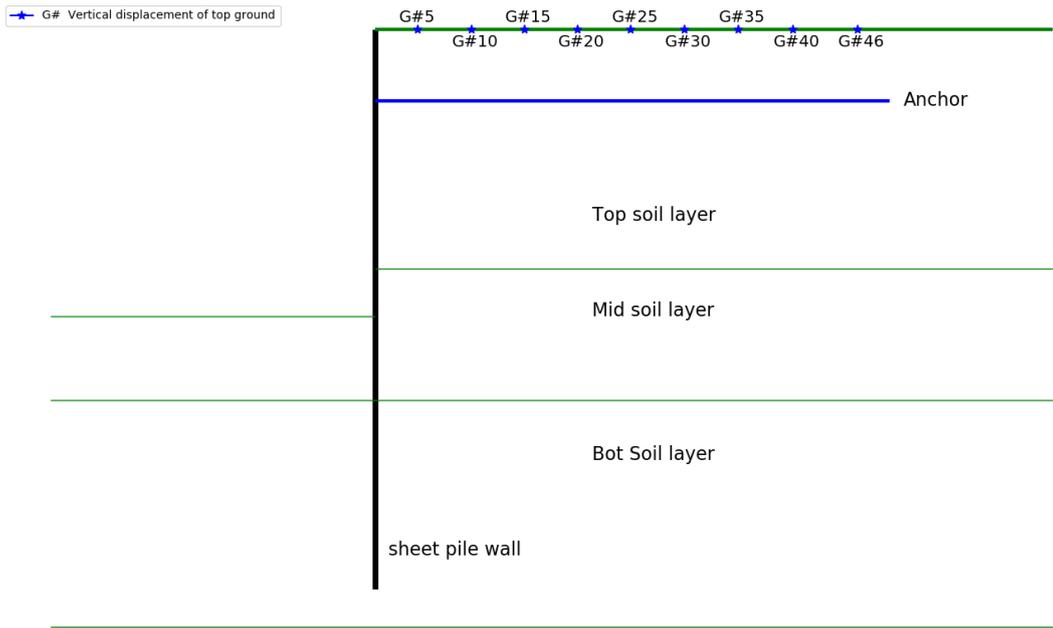


Figure 5.37: Sensor placement in Case 3-3.

Table 5.7: KLD in Case 3-3.

Sensor installation time		P4 _{start}	P5 _{start}	P6 _{start}	P7 _{start}
1 optimal sensor	Location	G#46	G#10	G#5	G#10
	KLD	2.66	1.45	1.44	1.34
	KLD (Normalized)	0.507	0.276	0.274	0.255
2 optimal sensor	Location	G#5, G#40	G#5, G#25	G#5, G#25	G#5, G#25
	KLD	3.82	1.99	1.93	1.87
	KLD (Normalized)	0.728	0.379	0.368	0.356
3 optimal sensor	Location	G#5, G#25, G#35	G#5, G#20, G#40	G#5, G#20, G#35	G#5, G#20, G#35
	KLD	4.28	2.37	2.24	2.17
	KLD (Normalized)	0.815	0.451	0.427	0.413
All sensors	Location	G#5, G#10, G#15, G#20, G#25, G#30, G#35, G#40, G#46			
	KLD	5.25	2.86	2.69	2.69
	KLD (Normalized)	1.000	0.545	0.512	0.512

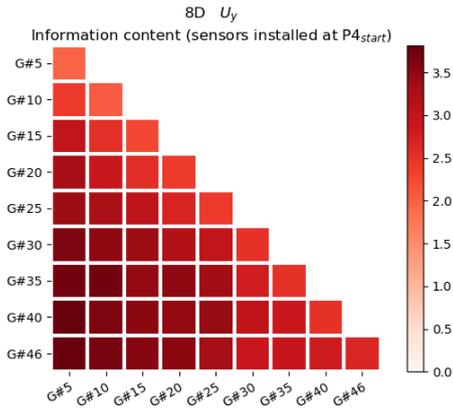


Figure 5.38: KLD matrix plot with sensors installed at $P4_{start}$.

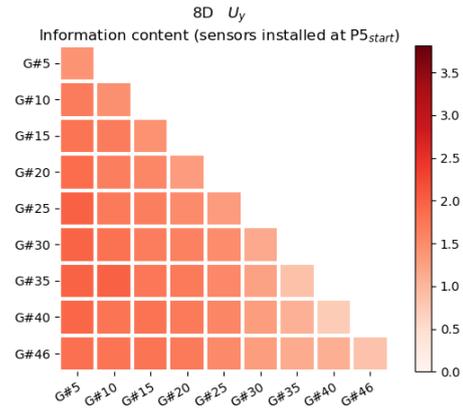


Figure 5.39: KLD matrix plot with sensors installed at $P5_{start}$.

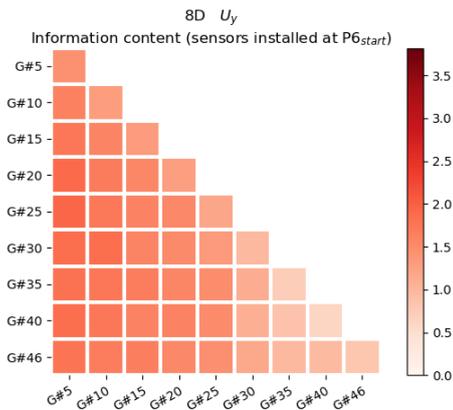


Figure 5.40: KLD matrix plot with sensors installed at $P6_{start}$.

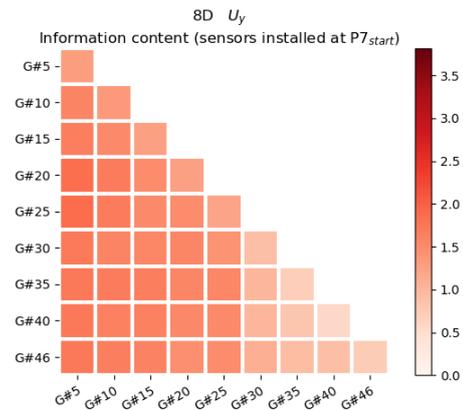


Figure 5.41: KLD matrix plot with sensors installed at $P7_{start}$.

From the summarized plot Figure 5.42, the main observations still hold the same:

- Three optimal sensors are sufficient whenever they are installed (they can carry 90% of the information compared of all 9 sensors) as the influence of correlation is small.
- The earlier the sensor installation, the more the carried information. However, it does make little difference whether the sensors are installed at $P5_{start}$, $P6_{start}$, $P7_{start}$, sensors installed at $P4_{start}$ could carry around 40% more information than those installed later. This can be verified from Figure 4.6 where the increment of vertical displacement in the P5 and P6 is tiny.
- Only E_{top} and C_{top} can be properly inferred even with the optimal sensor placement shown in Figure 5.43. In Case 3-4, this can be improved by adding more weekly correlated sensors (i.e. the optimal sensors in Case 3-1, Case 3-2 and Case3-3 as well as strain sensor in anchor will be combined).

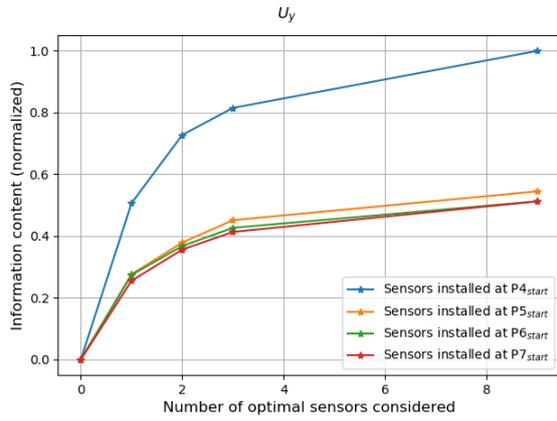


Figure 5.42: Variation of KLD in Case 3-3.

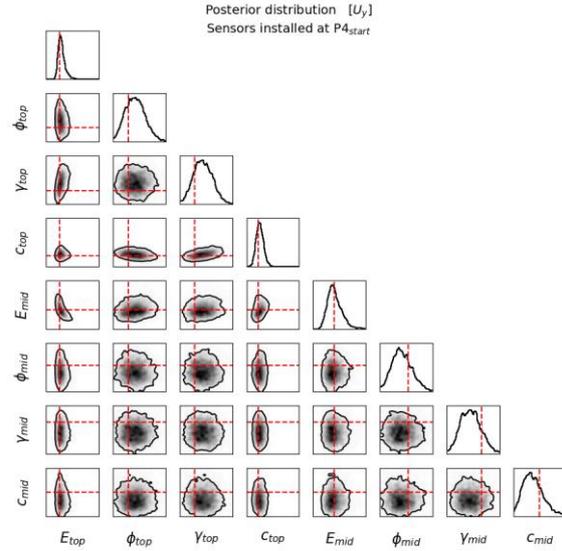


Figure 5.43: The most informative posterior in Case 3-3.

5.3.1.4 Case 3-4: Combination of all optimal sensors

The results of previous sub-sections show that none of the three sets of observations ε_{spw} , U_x or U_y could accurately estimate the eight soil parameters. Therefore, in this section we analyze what happens if a mixture of sensor types are used in the parameter estimation. For each sensor type, the three optimal sensors from the sections 5.3.1.1, 5.3.1.2 and 5.3.1.3 are selected, since three optimal sensors carry a comparable ($\sim 80\%$) information content as all the available sensors. Together with the single strain sensor at the anchor (ε_{anchor}), these in total ten sensors are used in the Bayesian analysis. The correlation between the same type of sensors is kept the same as done previously, while the correlation between different type of sensors is considered by setting $\hat{\rho} = 0.8$ as introduced in section 3.3.2.1. Since the optimal sensor placement varies with sensor installation phase, the corresponding configurations for each installation phase are displayed in Figure 5.44 ~ Figure 5.47. The KLD results obtained from the inference are collected and compared with earlier results in Table 5.8.

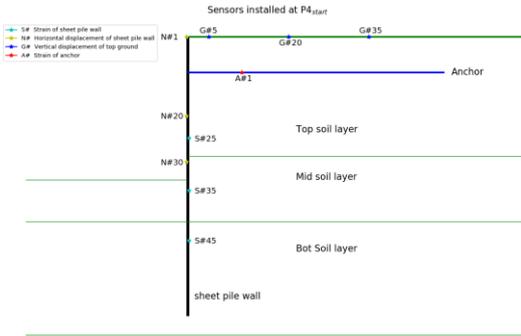


Figure 5.44: Optimal sensor placement at $P4_{start}$ in Case 3-4.

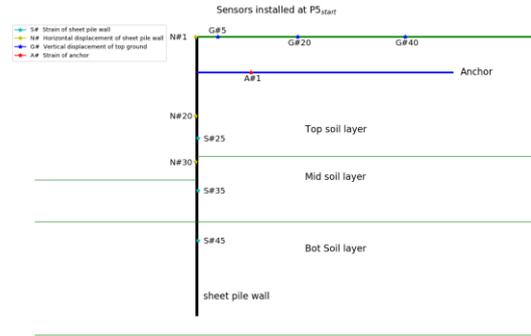


Figure 5.45: Optimal sensor placement at $P5_{start}$ in Case 3-4.

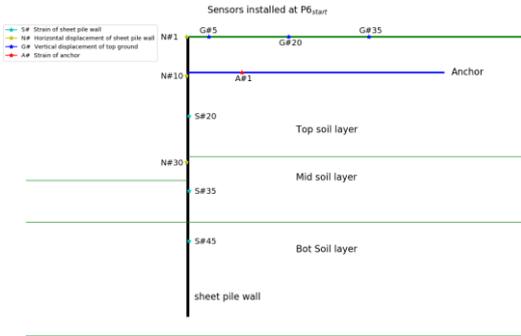


Figure 5.46: Optimal sensor placement at $P6_{start}$ in Case 3-4.

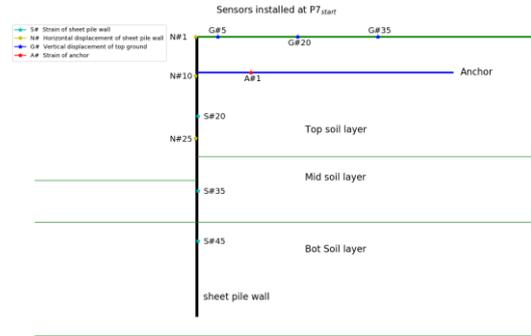


Figure 5.47: Optimal sensor placement at $P7_{start}$ in Case 3-4.

Table 5.8: KLD in Case 3-4.

Sensor installation time		$P4_{start}$	$P5_{start}$	$P6_{start}$	$P7_{start}$
ϵ_{spw}	3 optimal sensors	N#25, N#35, N#45	N#25, N#35, N#45	N#20, N#35, N#45	N#20, N#35, N#45
	KLD	4.06	3.91	2.50	1.85
	KLD (Normalized)	0.560	0.539	0.345	0.255
U_x	3 optimal sensors	N#1, N#20, N#30	N#1, N#20, N#30	N#1, N#10, N#30	N#1, N#10, N#25
	KLD	2.46	1.74	1.10	0.91
	KLD (Normalized)	0.339	0.240	0.152	0.126
U_y	3 optimal sensors	G#5, G#25, G#35	G#5, G#20, G#40	G#5, G#20, G#35	G#5, G#20, G#35
	KLD	4.28	2.37	2.24	2.17
	KLD (Normalized)	0.590	0.327	0.309	0.299
ϵ_{anchor}	Only 1 sensor	Not considered alone	Not considered alone	Not considered alone	Not considered alone
Combination of all 10 sensors	KLD	7.25	5.42	4.35	3.79
	KLD (Normalized)	1.000	0.748	0.600	0.523

The results in Table 5.8 are also plotted in Figure 5.48, leading to the following observations:

- As expected, the general rule still holds for the combination of ten sensors: the earlier the sensor installation phase, the more the conveyed information. And each delay of installation phase (i.e. from $P4_{start}$ to $P5_{start}$, from $P5_{start}$ to $P6_{start}$, from $P6_{start}$ to $P7_{start}$)

will respectively lose around 25%, 15% and 10% information (given as a percentage of the information that is gained with sensors installed at $P4_{start}$).

- U_x carries the least information (i.e. around 35% of the information conveyed by the combined ten sensors);
- ε_{spw} and U_y are much better than U_x , they in average convey around 60% of the information carried by the combined sensors. Albeit their performance differs much if they are installed at $P5_{start}$, in which case ε_{spw} conveys 20% more information than U_y . This large difference is due to that U_y occurred in P4 is the largest portion during the whole construction phase and if sensors are install at $P5_{start}$, the corresponding information is not collected.
- Even though the combination of all ten sensors can provide much more information regardless of sensor installation phase, only E_{top} , C_{top} and E_{mid} can be relatively well identified, while the marginal distributions of the remaining parameters are still more resembling their prior distributions as shown in Figure 5.49, which is based on sensors installed at $P4_{start}$.

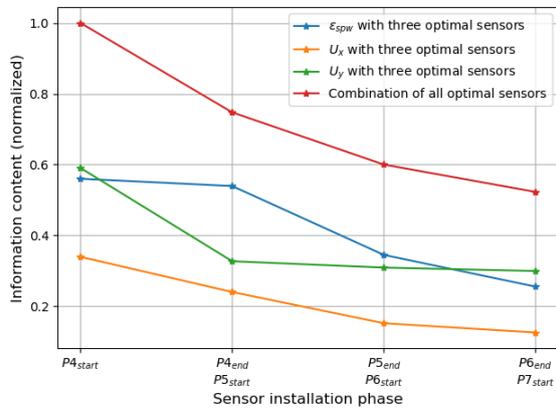


Figure 5.48: Variation of KLD in Case 3-4.

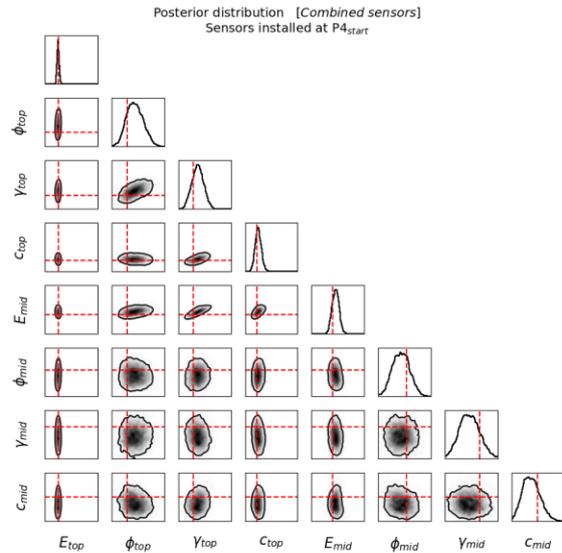


Figure 5.49: The most informative posterior in Case 3-4.

This non-identifiability is determined by the mechanical behavior of the structure, because the unidentified parameters play limited role in the structural response. To support this claim, the responses corresponding to these ten optimal sensors installed at $P4_{start}$ are displayed in Figure 5.50, in which each row corresponds to one of the optimal sensors, and each column corresponds to a soil parameter. In each subplot only one parameter (the soil parameter) is varied while the remaining parameters are fixed at the “real” values indicated by the red dashed

lines. From this figure, it can be clearly seen that changing ϕ_{top} , ϕ_{mid} , γ_{mid} , and C_{mid} hardly changes the responses (2nd, 6th, 7th, and 8th columns on the figure).

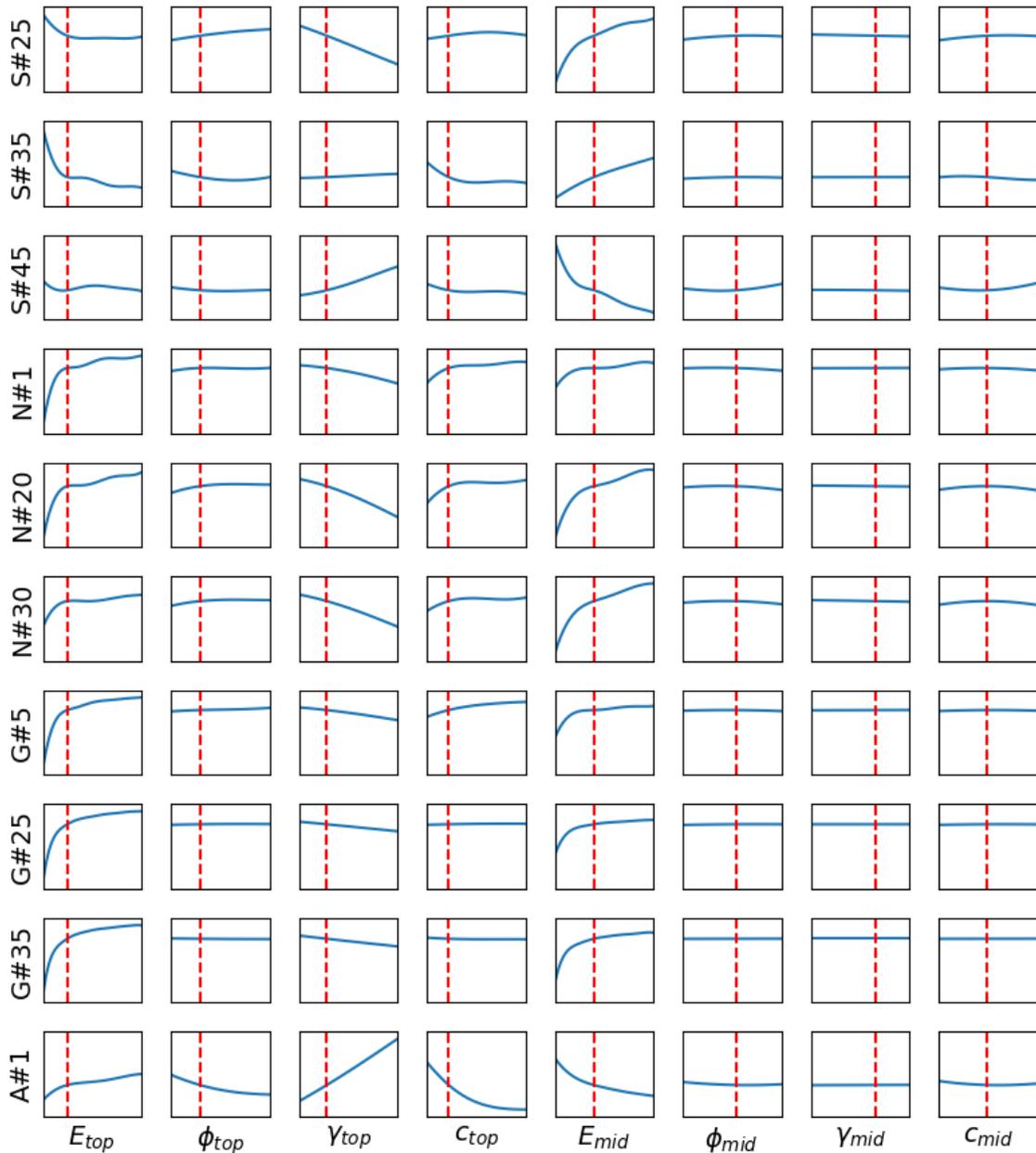


Figure 5.50: Response with sensors installed at $P4_{start}$ in Case 3-4.

5.3.2 Reliability assessment

In this section, the influence of system identification on both the serviceability limit state and ultimate limit state of this sheet pile wall system is investigated. The posterior distributions of system identification are based on the data from the combined optimal sensors in section 5.3.1.4, and these posteriors are used as inputs for the reliability analyses. The exceedance of horizontal displacement of sheet pile wall is selected as the serviceability limit state as in section 5.1.2

and 5.2.4. The exceedance of strain/stress limit of the sheet pile wall or/and anchor is considered as the ultimate limit state, in which the sheet pile wall and anchor are firstly considered separately and then together as a system. Since cumulative responses are required for the reliability calculation, the surrogate models of the responses at sensors corresponding to P4_{start} are reused in the assessment.

5.3.2.1 Serviceability limit state (SLS)

The serviceability limit state concerns the exceedance of the horizontal displacement threshold of the sheet pile wall. The same performance function is used as in section 5.1.1 (see equation (5.1)). Reliability analyses are completed using both the prior and posteriors obtained in Case 3-4 (i.e. four posteriors in total corresponding to combined sensors installed at P4_{start}, P5_{start}, P6_{start} and P7_{start}). The calculated failure probabilities are collected in Table 5.9 and plotted in Figure 5.51. Notice the increase in prior failure probability: from 9.72×10^{-2} (Table 5.2) to 1.50×10^{-1} (Table 5.9). The increase is solely caused by the increasing dimension of the probability space and in turn the introduction of additional uncertainties.

Table 5.9: Failure probability in SLS.

Sensor installation time	P4 _{start}	P5 _{start}	P6 _{start}	P7 _{start}
Prior	1.50×10^{-1}	1.50×10^{-1}	1.50×10^{-1}	1.50×10^{-1}
Posterior	$< 1.50 \times 10^{-300}$ *	2.47×10^{-157}	1.57×10^{-71}	3.31×10^{-38}

* The failure probability is so small that nested sampling with the used settings could not estimate it reliably.

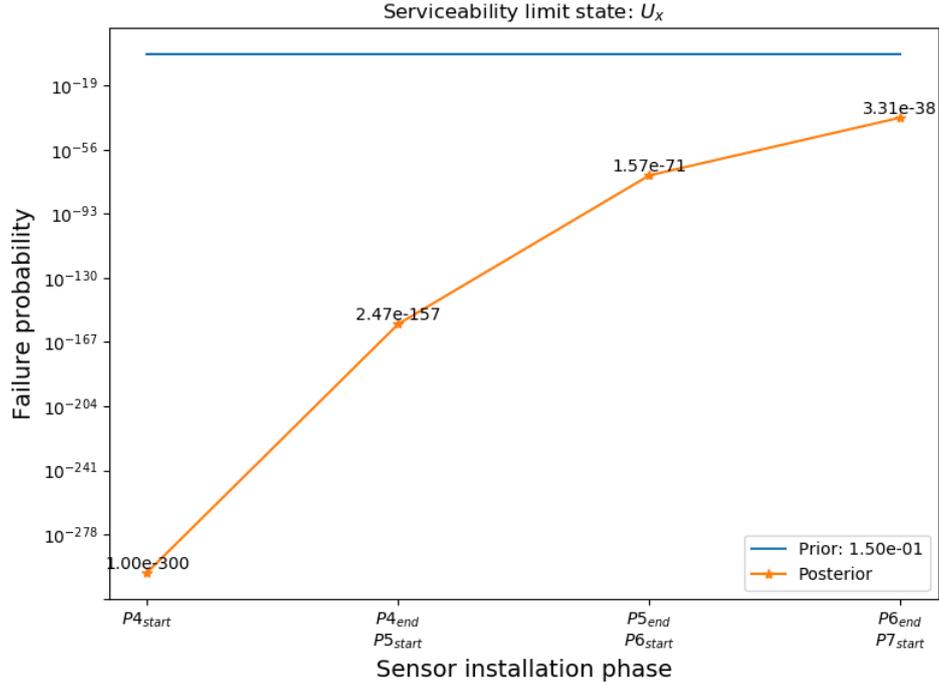


Figure 5.51: Prior and posterior failure probabilities in SLS for different sensor installation times.

From this figure, the following observations can be made:

- Even if the same \hat{U}_x is applied to the limit state function in this subcase, the prior failure probability is increased to 15% compared with that in Subcase 1 and Subcase 2 in which it is around 10%. This is because with the increase of dimensionality, the volume of random variables increases, leading to larger uncertainty about the variable space.
- The failure probability is reduced dramatically from prior (i.e. 15%) to posterior. Even the largest value corresponding to sensors at $P7_{start}$ is a negligible value (i.e. 3.31×10^{-38}). This comparison can be seen straightforwardly from Figure 5.52 and Figure 5.56, in which the red zones indicate the projection of the failure region. The huge difference of failure probability is due to that most of the failure projections fall within the 90% credible region in prior distribution (see E_{top} & E_{mid} square in Figure 5.52), while the failure projections are far away from the 90% credible region in posterior distribution (by checking all the squares in Figure 5.56);
- There is a sharp increase in failure probability with the delay of sensor installation time, since the posterior is more and more dispersed and the 90% credible regions are getting closer to the failure projections.

- The largest failure probability in posterior is still tiny (has the magnitude of 1×10^{-38}). This is determined by the mechanical behavior of this structure and the selected “real” values of soil properties, since the posteriors concentrate around the “real” values, based on which the calculated maximum of horizontal displacement of sheet pile wall (i.e. around $100mm$) is far below the threshold $\hat{U}_x = 180mm$.

Prior distribution

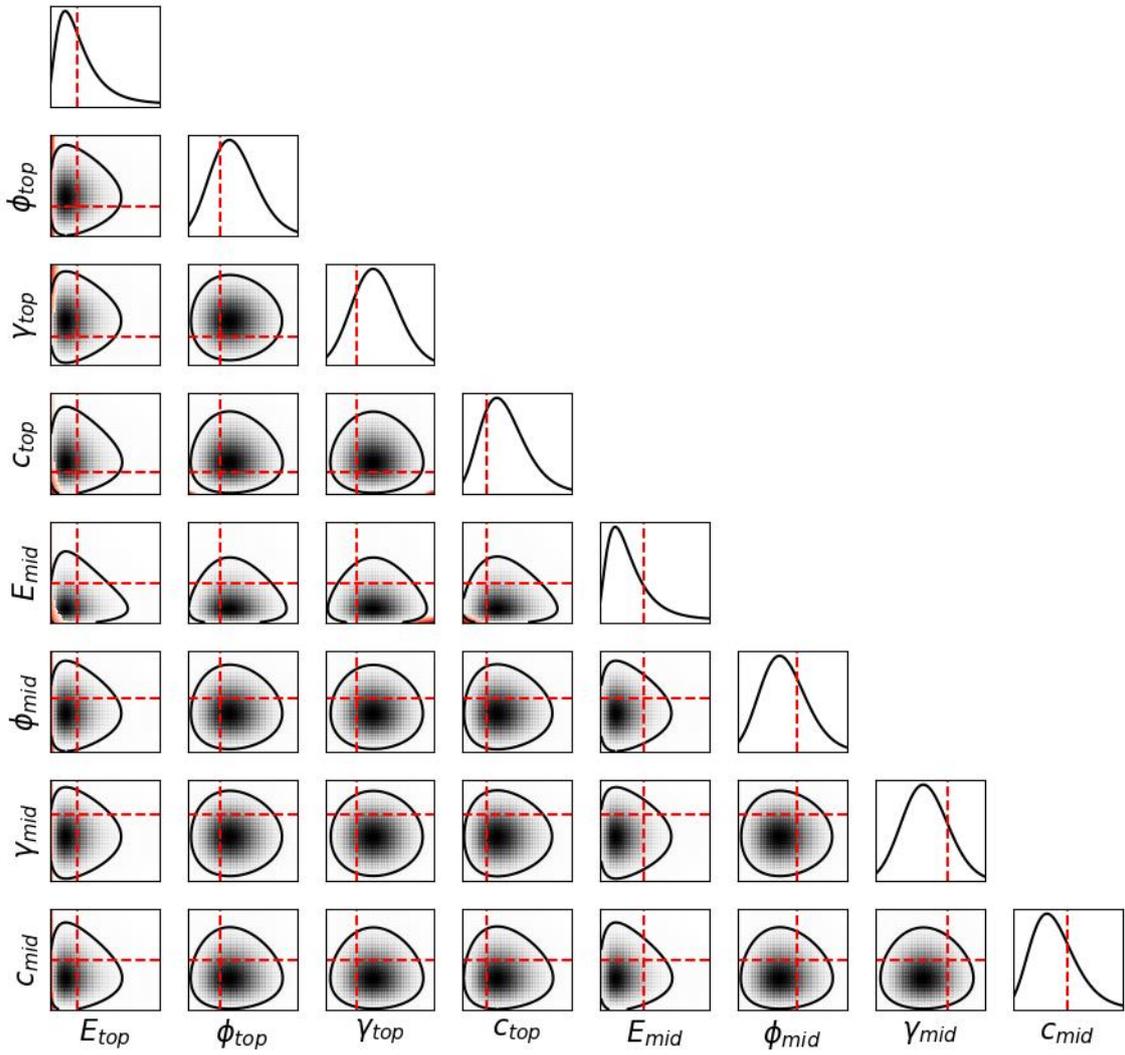


Figure 5.52: Projection of failure regions in prior.

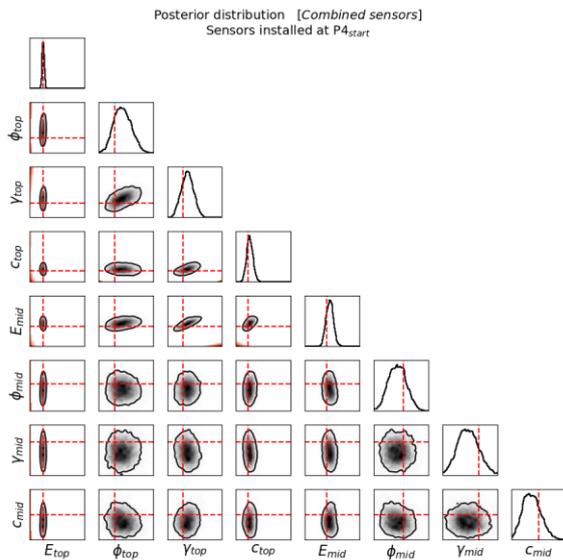


Figure 5.53: Projection of failure regions in posterior with sensors installed at $P4_{start}$.

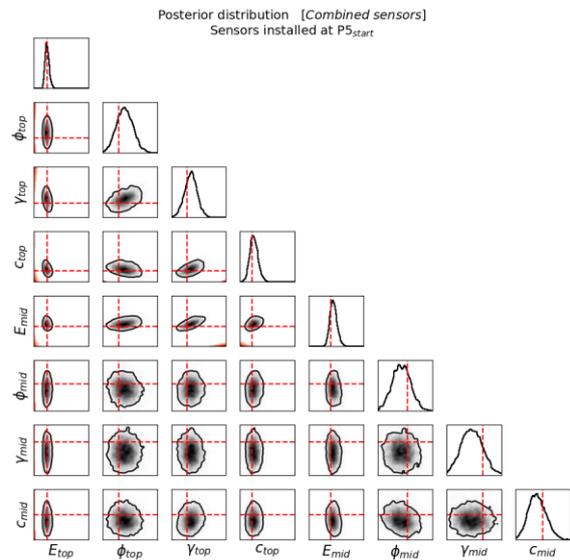


Figure 5.54: Projection of failure regions in posterior with sensors installed at $P5_{start}$.

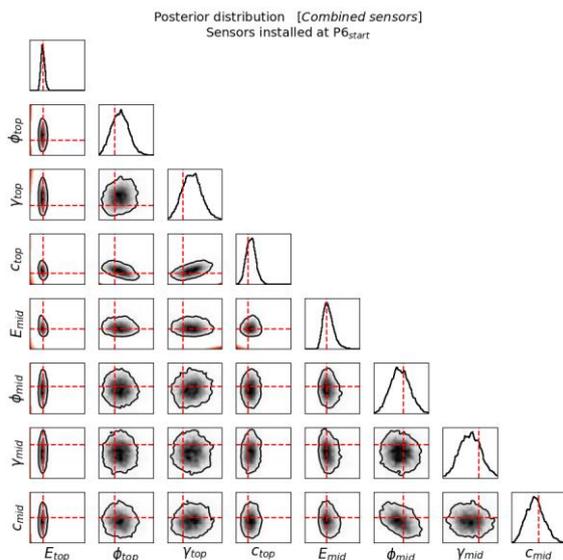


Figure 5.55: Projection of failure regions in posterior with sensors installed at $P6_{start}$.

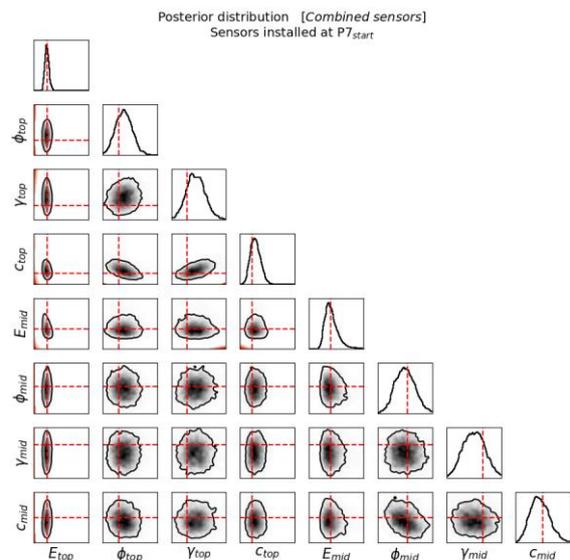


Figure 5.56: Projection of failure regions in posterior with sensors installed at $P7_{start}$.

5.3.2.2 Ultimate limit state (ULS)

Sheet pile wall failure

Firstly, the exceedance of the yield stress of sheet pile wall is considered. The stresses are caused by the bending moments and axial forces. Since the strain of the sheet pile wall is already available and surrogated in system identification, we can reuse the responses to obtain

stresses by simply multiplying them by the elastic modulus of steel (i.e. $2.1 \times 10^5 N/mm^2$). Then the limit state function is given:

$$g_{\text{spw}}(\boldsymbol{\theta}) = \hat{\sigma}_{\text{spw}} - \max(\sigma_{i,\text{spw}}(\boldsymbol{\theta})) \quad (5.2)$$

Where

$\hat{\sigma}_{\text{spw}}$ Yield strength of sheet pile wall, taken as a random variable with lognormal distribution;

$\sigma_{i,\text{spw}}(\boldsymbol{\theta})$ Stress of sheet pile wall at S# i , $i = 20, 25, 30, 35, 40, 45$ as shown in Figure 5.22.

These positions (i.e. S# i) are selected in reliability analysis as only limited strain is expected at the top part and bottom part of sheet pile wall (see Figure 4.4), it is hence efficient to discard the two portions to reduce computational time. The coefficient of variation (COV) of $\hat{\sigma}_{\text{spw}}$ is set as 0.07 according to (Teixeira et al., 2015), and the mean is set as $325 N/mm^2$ to obtain a high failure probability (i.e. $\sim 1\%$) in prior distributions.

Since in the performance function, the resistance (i.e. $\hat{\sigma}_{\text{spw}}$) is now a random variable instead of a constant as in serviceability limit state, the projection of failure regions is not available, and equation (3.57) is applied. The limit state function is applied to the four posteriors obtained in Case 3-4 (see section 5.3.1.4) and the results are collected in Table 5.10.

Table 5.10: Failure probability in sheet pile wall failure.

Sensor installation time	P4 _{start}	P5 _{start}	P6 _{start}	P7 _{start}
Prior	1.11×10^{-2}	1.11×10^{-2}	1.11×10^{-2}	1.11×10^{-2}
Posterior	3.60×10^{-26}	5.14×10^{-25}	1.18×10^{-13}	3.29×10^{-11}

The results from Table 5.10 are also visualized in Figure 5.57, from which the following observations can be made:

- The reduction of failure probability from prior to posterior after system identification is still dramatic (the smallest $P_{f,\text{prior}}$ to $P_{f,\text{posterior}}$ ratio is about 10^9 when sensors are installed at P7_{start});
- The obtained failure probability is getting larger and larger with the delay of sensor installation phase, particularly from P5_{start} to P6_{start} the ratio of the calculated failure probabilities is 10^{12} . This is consistent with the trend of KLD value of ε_{spw} in Figure

5.48, since in P5 the excavation of top and mid soil layer (see section 4.2) will lead to large increase of strain/stress of sheet pile wall (see Figure 4.4), if sensors are installed after this phase, this large strain gain will not be collected, leading to a more dispersed posterior and a larger failure probability.

- The largest failure probability (i.e. sensors are installed at P7_{start}) in posterior is still small (has the magnitude of 1×10^{-11}). This is likely caused by the mechanical behavior of this structure, the considerable reduction in uncertainty in the posterior, and the selected “real” values of soil properties as discussed in section 5.3.2.1. The maximum of stress of sheet pile wall calculated using the “real” values (i.e. around $175 N / mm^2$) is far below the threshold (mean of $\hat{\sigma}_{spw} = 325 N / mm^2$).

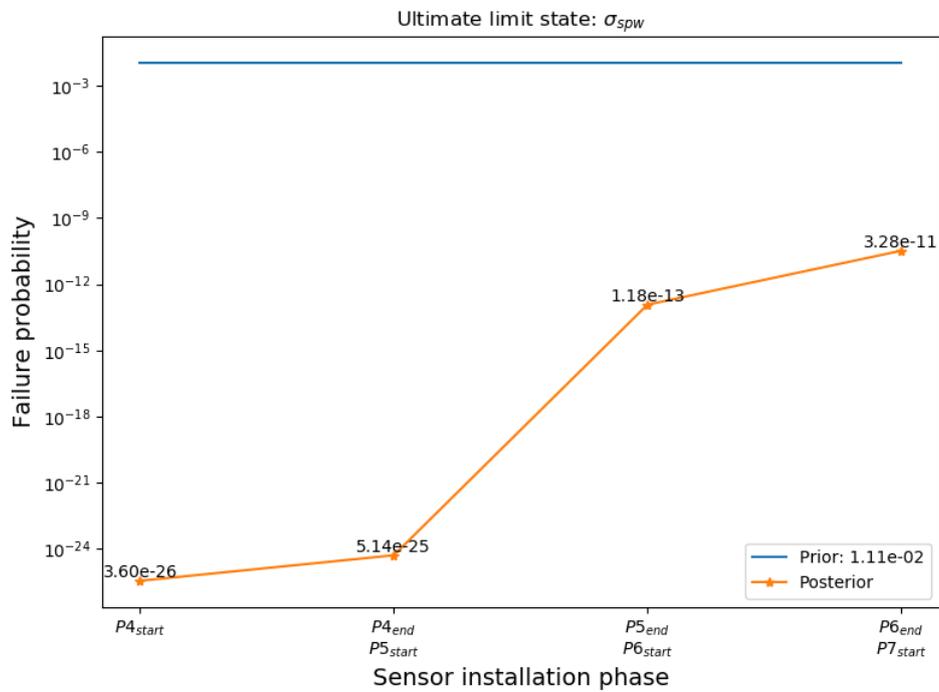


Figure 5.57: Showcase of failure probability in sheet pile wall failure.

Anchor failure

The anchorage failure is defined as the exceedance of the maximum stress of the anchorage’s yield stress. Since the anchor is modeled as a truss element, only an axial force contributes to the stress and the stress is considered to be a constant along the anchor. Then the limit state function is defined as:

$$g_{\text{anchor}}(\boldsymbol{\theta}) = \hat{\sigma}_{\text{anchor}} - \sigma_{\text{anchor}}(\boldsymbol{\theta}) \quad (5.3)$$

where

$\hat{\sigma}_{\text{anchor}}$ yield strength of anchor, taken as a random variable with lognormal distribution;
 $\sigma_{\text{anchor}}(\theta)$ stress of anchor.

The coefficient of variation (COV) of $\hat{\sigma}_{\text{anchor}}$ is set as 0.07 according to (Teixeira et al., 2015), and the mean is set as $120N/mm^2$ to obtain a high failure probability (i.e. ~1%) in prior distributions. Then the limit state function is applied to the four posteriors obtained in Case 3-4 (see section 5.3.1.4) and the results are collected in Table 5.11.

Table 5.11: Failure probability in anchor failure.

Sensor installation time	P4 _{start}	P5 _{start}	P6 _{start}	P7 _{start}
Prior	1.25×10^{-2}	1.25×10^{-2}	1.25×10^{-2}	1.25×10^{-2}
Posterior	3.95×10^{-10}	6.88×10^{-10}	1.25×10^{-6}	2.35×10^{-6}

The results in Table 5.11 are displayed in Figure 5.58, based on these results similar observations can be made as in the sheet pile wall failure:

- The variation of failure probability from prior to posterior after system identification is huge (the minimum of ratio is 10^4 when sensors are installed at P7_{start}).
- The obtained failure probability is getting larger and larger with the delay of sensor installation phase, particularly from P5_{start} to P6_{start} the ratio of the calculated failure probabilities is 10^4 due to the same reason as explained in sheet pile wall failure.
- The largest failure probability (i.e. sensors are installed at P7_{start}) in posterior is still small (has the magnitude of 1×10^{-6}) as explained in sheet pile wall failure: the stress of anchor calculated using the “real” values (i.e. around $75N/mm^2$) is far below the threshold (mean of $\hat{\sigma}_{\text{anchor}} = 120N/mm^2$).

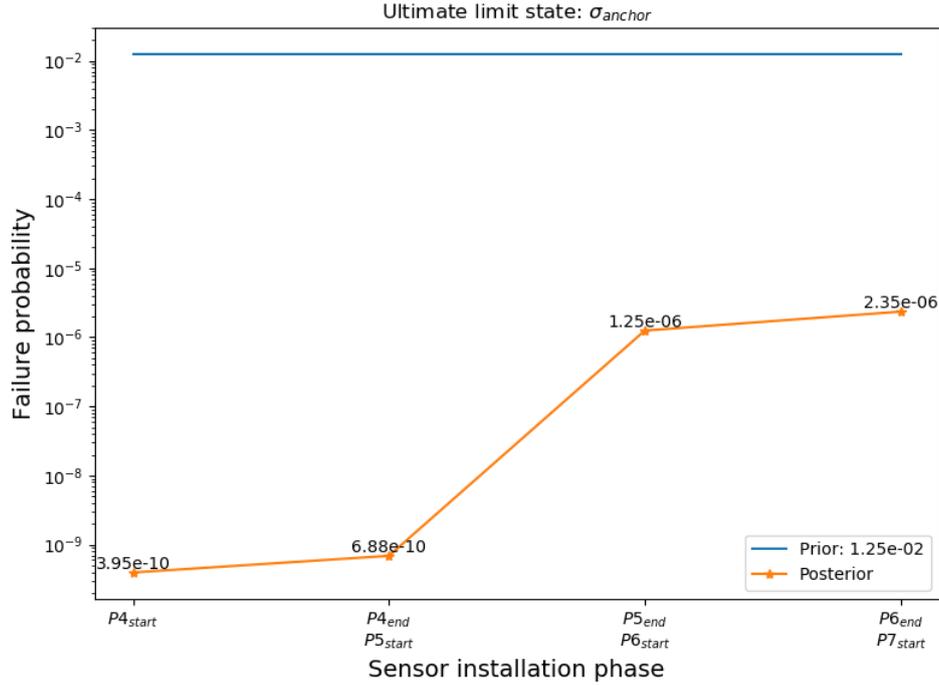


Figure 5.58: Showcase of failure probability in anchor failure.

Structural failure

In this section, the failure of the entire system is considered, combining sheet pile wall failure and anchor failure which are assumed to form a series system. The performance function is the following:

$$\begin{aligned}
 g_{\text{structure}}(\boldsymbol{\theta}) &= (\hat{\sigma}_{\text{spw}} - \max(\sigma_{i,\text{spw}}(\boldsymbol{\theta}))) \cup (\hat{\sigma}_{\text{anchor}} - \sigma_{\text{anchor}}(\boldsymbol{\theta})) \\
 &= \min(\hat{\sigma}_{\text{spw}} - \max(\sigma_{i,\text{spw}}(\boldsymbol{\theta})), \hat{\sigma}_{\text{anchor}} - \sigma_{\text{anchor}}(\boldsymbol{\theta})) \\
 &= \min(g_{\text{spw}}(\boldsymbol{\theta}), g_{\text{anchor}}(\boldsymbol{\theta}))
 \end{aligned} \tag{5.4}$$

The failure probability is calculated by the following equation:

$$\begin{aligned}
 P_f &= P[(g_{\text{spw}}(\boldsymbol{\theta}) < 0) \cup (g_{\text{anchor}}(\boldsymbol{\theta}) < 0)] \\
 &= P(g_{\text{spw}}(\boldsymbol{\theta}) < 0) + P(g_{\text{anchor}}(\boldsymbol{\theta}) < 0) - P[(g_{\text{spw}}(\boldsymbol{\theta}) < 0) \cap (g_{\text{anchor}}(\boldsymbol{\theta}) < 0)] \\
 &= P(g_{\text{spw}}(\boldsymbol{\theta}) < 0) + P(g_{\text{anchor}}(\boldsymbol{\theta}) < 0) - P(g_{\text{spw}}(\boldsymbol{\theta}) < 0) \times P(g_{\text{anchor}}(\boldsymbol{\theta}) < 0)
 \end{aligned} \tag{5.5}$$

It is crucial to mention that the last line in equation (5.5) holds only true if $g_{\text{spw}}(\boldsymbol{\theta}) < 0$ and $g_{\text{anchor}}(\boldsymbol{\theta}) < 0$ are independent. Since the failure of sheet pile wall and anchor are usually not independent, the result obtained using equation (5.5) is more conservative (i.e. larger failure probability than the exact solution).

The distributions of $\hat{\sigma}_{\text{spw}}$ and $\hat{\sigma}_{\text{anchor}}$ are set the same as in sheet pile wall failure and anchor failure. The prior and posterior (i.e. obtained in Case 3-4) failure probabilities are given in Table 5.12 and plotted in Figure 5.59. The obtained prior system failure probability is larger than that of the prior component failures separately (sheet pile wall failure and anchor failure). This is reasonable due to the expansion of failure domain. The posterior system failure probabilities are very close to those of the anchor failure alone. This indicates that the anchor failure is the dominating failure mode in the sheet pile wall system and thus the same observations due to the same reasons can be made as for the anchor failure alone.

Table 5.12: Failure probability in structural failure.

Sensor installation time	P4 _{start}	P5 _{start}	P6 _{start}	P7 _{start}
Prior	1.65×10^{-2}	1.65×10^{-2}	1.65×10^{-2}	1.65×10^{-2}
Posterior	3.95×10^{-10}	6.88×10^{-10}	1.25×10^{-6}	2.35×10^{-6}

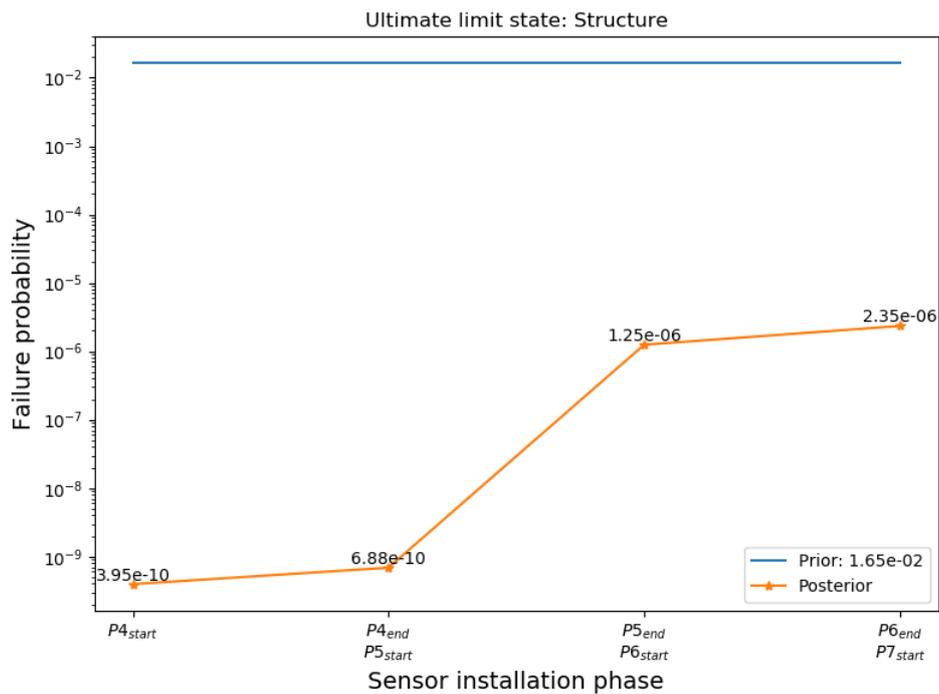


Figure 5.59: Showcase of failure probability in structural failure.

5.4 Discussion

The dimensionality of the probabilistic space has a large influence on system identification as well as reliability assessment. If we zoom into the marginal posterior distribution of E_{top} and E_{mid} obtained from Subcase 1, Subcase 2 and Case 3-1 (see Figure 5.3, Figure 5.7 and Figure 5.29) which are all based on 6 strain sensors installed at $P4_{\text{start}}$, it can be seen that both E_{top} and E_{mid} can be well identified as they are concentrated around the “real” value and have a much smaller COV compared with the prior. However, the COV in these three subcases increase gradually. The increment is 22%, 25% for E_{top} and 19%, 65% for E_{mid} when moving from 2D to 4D and from 2D to 8D respectively (all percentages are calculated in respect of the 2D case). This is because with the increase of dimensionality, the number of parameter combinations that match with the measured responses increase as well and thus the uncertainty of the variable space increases. The quantitative values of 90% ACR in Table 5.5 also support this observation, in which the 90% ACR of $E_{\text{top}} - E_{\text{mid}}$ in posterior in Case 3-1 (i.e. 8D problem) is almost 5 times larger than that in Subcase 1 (i.e. 2D problem). This situation can be improved a bit by adding more sensors with no/weak correlation. For example, C_{top} is not identifiable at all with only ε_{spw} sensors as shown in Figure 5.28, but it can be well inferred with combined sensors as shown in Figure 5.49. Due to the same reason, when the same serviceability limit state function is applied to 2D, 4D, and 8D posterior, the resulting failure probability increases (i.e. namely 5.11×10^{-88} , 2.43×10^{-80} , and 7.39×10^{-56}).

The construction stages also play an essential role in system identification and reliability assessment. With each delay of sensor installation time, there will be a 10% ~ 25% decrease of gained information compared with that gained from sensors installed at $P4_{\text{start}}$, leading to a dispersed posterior and thus a larger failure probability in the reliability analysis (both SLS and ULS). This decreasing trend holds true regardless of the type of sensors and number of sensors. There might be a particular construction phase after which the installed sensors carry much less information. For example, 20% more information can be gained by strain sensors if they are installed at $P5_{\text{start}}$ instead of $P6_{\text{start}}$, because large strain is expected in P5 caused by excavation of soil but is not recorded. This also explains why there is a jump of ultimate limit state from $P5_{\text{start}}$ to $P6_{\text{start}}$ shown in Figure 5.57, Figure 5.58 and Figure 5.59.

In both SLS and ULS, the gain from monitoring is huge: there is a dramatic decrease of failure probability in posterior compared with that in prior. One reason is that the responses calculated

by the selected “real” values are far below the threshold, while another is that the uncertainty of the selected prior (i.e. coefficient of variation) is much larger compared with the posterior.

6 Conclusions and recommendations

6.1 Answers to main research question

In the context of hydraulic structures, this thesis aims at finding a feasible way to

- 1) identify soil parameters based on a monitoring system considering all relevant sources of uncertainty, correlation between measurements, and relatively high dimensionality;
- 2) investigate the influence of the inclusion of monitoring data (updating the prior to posterior) on structural reliability;
- 3) investigate the influence of construction stages on both parameter identification and reliability assessment.

To answer the main research questions:

What is the effect of monitoring hydraulic structures on structural reliability? What is the influence of starting the monitoring at different construction stages?

A series of analyses are performed considering a single-anchor sheet pile wall system. Firstly, up to eight unknown soil parameters are identified by a Bayesian approach considering measurement uncertainty, physical model uncertainty, and surrogate model uncertainty as well as their correlation. This identification process is performed four times to obtain four sets of soil parameter posterior distributions, considering multiple types and number of sensors and that sensors could be installed at four different construction stages to collect structural responses. In the subsequent reliability assessment, the obtained four sets of posteriors serve as the inputs to calculate the failure probability of the structure considering both serviceability limit state and ultimate limit state. The following conclusions can be drawn related to the main research questions:

- The monitoring of hydraulic structures (system identification) can largely decrease the failure probability (P_f) in both SLS and ULS. Particularly, in this work the smallest $P_{f,prior}$ to $P_{f,posterior}$ ratios (considering different construction stages) are 1×10^{37} in SLS and 1×10^4 in ULS; while the largest ratios are 1×10^{299} in SLS and 1×10^{24} in ULS.
- The monitoring should start at as early construction stage as possible to collect sufficient information, which is beneficial for both parameter identification and reliability assessment. With each delay of sensor installation time, the gained

information will decrease, leading to a more dispersed posterior and thus an increase of failure probability. For example in this case study, the decrement of KLD value can be 30% (see Case 3-4 in section 5.3.1.4 in which sensors are installed from P4_{start} to P5_{start}), while the $P_{f,prior}$ to $P_{f,posterior}$ ratios 1×10^{12} (see sheet pile wall failure in section 5.3.2.2). This decreasing and increasing trend holds true regardless of dimensionality, type of sensors and number of sensors.

Although some explanations are given in this thesis for the calculated unexpectedly large ratio of prior and posterior failure probabilities, the explanations are not deemed to be sufficiently detailed and deep. Therefore, further research is needed to better understand its source and gain further confidence in the conclusions: "The weight of evidence for an extraordinary claim must be proportioned to its strangeness."

6.2 Answers to sub research questions

In practical implementation, the sub questions raised from the main research questions as introduced in section 1.3 can also be answered.

1. What type of sensors can be and should be used?

In this work, we consider four types of observations to give some insight, namely strain of sheet pile wall (ε_{spw}), horizontal displacement of sheet pile wall (U_x), vertical displacement of top ground (U_y) and strain of anchor (ε_{anchor}). Based on their conveyed information, we have this order: $\varepsilon_{spw} \approx U_y \approx 1.5 \times U_x$ (ε_{anchor} is not considered alone). However, none of them could solely and properly identify the soil parameters. Since their combination outperforms any one of them (i.e. 40% more information conveyed), a combination of sensor types is preferred.

2. How many sensors are needed, where to put these sensors and when to install them (during construction or afterwards)?

Generally, the more uncorrelated sensors, the better the inference result, which however in reality is usually restricted by the budget considering e.g. cost of sensors, cost of installation, cost of maintenance, etc. The sensor locations vary with different construction stages but should be installed as early as possible. In this case study, from the result in Subcase 2 and Subcase 3, three optimal sensors (here the optimal means the minimum number of sensors) for each observations (i.e. ε_{spw} , U_x , and U_y) are sufficient to do parameter identification (only one strain sensor is needed on the anchor as the axial force is assumed constant along the length),

as they can provide around 90% of information carried by full sensor configuration. The locations of these optimal sensors vary with observation types and sensor installation stage, and all of them should be installed at the earliest construction stage to carry the most information (e.g. sensors installed at $P4_{\text{start}}$ can convey double information compared with sensors installed at $P7_{\text{start}}$).

3. Which Bayesian computational method fits best in this context that is characterized by computationally expensive likelihood functions considering correlation, high dimensionality (i.e. more parameters are considered simultaneously) and multimodal and/or degenerate posteriors?

MultiNest outperforms the traditional MCMC families considering the above requirements (the computationally expensive likelihood is coped with in sub question 5). Since MultiNest does not rely on random walk (e.g. Metropolis Hastings MCMC) which has high rejection rate or gradient of posterior (e.g. Hybrid MCMC) which slows down the calculation, but compresses the multidimensional integral of evidence into a one-dimensional integral, the calculation is efficient and a large amount of computation time can be saved. Due to the same reason, there is little trouble to deal with high dimensionality and multimodal/degenerate posteriors. Besides the posterior, the sample weight as a by-product in MultiNest can be used straightforwardly in reliability analysis without additional efforts. Based on the performance of MultiNest in our work, it can be applied to other hydraulic structures and even other civil structures as well.

4. How to quantitatively evaluate the information conveyed by sensors?

In this work, both 90% credible region and Kullback–Leibler (KL) divergence are used to evaluate the sensor based information. The former, which combines the information from both prior and sensors, is not quantitatively calculated but only plotted to intuitively assess how concentrated a posterior is, while the latter is numerically computed to quantitatively indicate the gained information solely from the sensors. With KL divergence, we can quantitatively evaluate and compare the performance of different sensors and determine the location and the number of optimal sensors.

5. How to construct an accurate surrogate model of the physical model in a reasonable computation time?

In this work, to deal with the computationally expensive likelihood function in sub question 3, surrogated models are constructed to replace the physical model (FE model). Gaussian process

regression (GPR) with anisotropic RBF kernel and white kernel is proved to perform well. An adaptive fitting procedure is used that is based on the prediction variance in GPR. The white kernel can protect against overfitting. These two components efficiently facilitate the accuracy of the surrogate models. The resulted surrogate models can be used not only in system identification but also in reliability assessment.

As a general conclusion: I demonstrated that the system identification and reliability assessment of realistic hydraulic structures (complex physical model and relatively high (up to 8) dimensional probabilistic space, large number of sensors (up to 10)) is feasible and I proposed, implemented, and tested all the components for such an analysis.

The above conclusions focus on the practical relevance; although, while seeking answers to the main and sub research questions to following noteworthy contributions are made to the system identification literature (a brief recapitulation of section 1.5 for convenience). To my knowledge those ended with star (*) are believed to be the first time in the open literature:

- Combination of system identification and structural reliability for hydraulic structures.*
- Analysis of the impact of construction stages on SHM of hydraulic structures.
- Application of nested sampling for SHM based inference and reliability analysis.*
- Proposal of a novel surrogating approach (active learning Gaussian process regression).
- Proposal for a general mathematical model that encompasses measurement uncertainties, physical model uncertainties, and surrogate model uncertainties.
- The quantification and inclusion of the surrogate model uncertainties in the system identification of hydraulic structures.*
- Implementation of a custom, Python-based program related to programmatic control of Plaxis, inference, reliability assessment (see Annex A to find the used Python packages).

Although there are many limitations of this work, e.g. all the quantitative results are particular to the selected single anchored sheet pile wall and synthetic observations are used (see the detailed list in section 1.4), the obtained results are promising and encouraging to continue this line of research.

6.3 Recommendations

Given the single case study considered in this work, it is too early to make general recommendations or recommendations for practical applications, hence I solely focus on recommendations for future research.

Future analysis could investigate:

- The influence of different priors (e.g. different COV of prior distribution);
- The influence of different “real” values of soil properties on reliability analysis;
- The influence of replacing the model that generates the synthetic data with a more sophisticated model (e.g. 3D and/or hardening soil model) while keeping the current model for system identification and reliability analysis;
- The individual contribution of three uncertainty components, namely measurement uncertainty, physical model uncertainty and surrogate model uncertainty;
- The influence of correlation length;
- The influence of more candidates of optimal sensors;
- The influence of dimensionality on the adaptive surrogating (e.g. when it becomes computationally expensive).

Besides, more interesting aspects can be studied, such as:

- What is the influence of sheet pile wall length on information content?
- What if loads are mainly from water fluctuation?
- What if sensors are installed at a fixed construction phase but data is collected at different phases? How to consider their correlation?

Furthermore, the investigation should be extended to additional structures such as different sheet pile walls, quay walls, etc. to be able to generalize the conclusions and in turns to derive practical recommendations.

References

- Arai, K., Ohta, H., & Kojima, K. (1987). Estimation of nonlinear constitutive parameters based on monitored movement of subsoil under consolidation. *SOILS AND FOUNDATIONS*, 27(1), 35-49. doi:10.3208/sandf1972.27.35
- Baily, E., Kieran, P., Goggins, J., & McCabe, B. A. (2014). *Real-time monitoring and performance of retaining structures*. Paper presented at the Civil Engineering Research in Ireland (CERI2014) conference.
- Balageas, D. J. S. h. m. (2006). Introduction to structural health monitoring. 13-43.
- Beck, J. L., & Au, S. K. (2002). Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation. *Journal of Engineering Mechanics-Asce*, 128(4), 380-391. doi:10.1061/(Asce)0733-9339
- Brinkgreve, R. B. J., Kumarswamy, S., Swolfs, W. M., Zampich, L., & Manoj, N. R. (2019). Plaxis manual.
- Capecchi, D., & Vestroni, F. (1999). Monitoring of structural systems by using frequency data. *Earthquake Engineering & Structural Dynamics*, 28(5), 447-461. doi:10.1002/(Sici)1096-9845(199905)28:5<447::Aid-Eqe812>3.3.Co;2-U
- Chen, B.-r., Zhao, H.-b., Ru, Z.-l., & Li, X. (2015). Probabilistic back analysis for geotechnical engineering based on Bayesian and support vector machine. *Journal of Central South University*, 22(12), 4778-4786. doi:10.1007/s11771-015-3029-1
- Ching, J. Y., & Chen, Y. C. (2007). Transitional markov chain monte carlo method for Bayesian model updating, model class selection, and model averaging. *Journal of Engineering Mechanics*, 133(7), 816-832. doi:10.1061/(Asce)0733-9399(2007)133:(816)
- Clayton, C. R. I., Woods, R. I., Bond, A. J., & Milititsky, J. (2014). *Earth Pressure and Earth-Retaining Structures* (Third edition ed.): CRC Press
- Collins, J. D., Hart, G. C., Hasselman, T. K., & Kennedy, B. (1974). Statistical Identification of Structures. *Aiaa Journal*, 12(2), 185-190. doi:10.2514/3.49190
- Dascotte, E., Strobbe, J., & Tygesen, U. T. (2013). *Continuous stress monitoring of large structures*. Paper presented at the Proceeding of 5th International Operational Modal Analysis Conference, Portugal.

- Deng, J., Lee, C. J. I. j. o. r. m., & sciences, m. (2001). Displacement back analysis for a steep slope at the Three Gorges Project site. *38(2)*, 259-268.
- Deng, J. H., & Lee, C. F. (2001). Displacement back analysis for a steep slope at the Three Gorges Project site. *International Journal of Rock Mechanics and Mining Sciences*, *38(2)*, 259-268. doi:[https://doi.org/10.1016/S1365-1609\(00\)00077-0](https://doi.org/10.1016/S1365-1609(00)00077-0)
- Diez, A., Khoa, N. L. D., Makki Alamdari, M., Wang, Y., Chen, F., & Runcie, P. (2016). A clustering approach for structural health monitoring on bridges. *Journal of Civil Structural Health Monitoring*, *6(3)*, 429-445. doi:10.1007/s13349-016-0160-0
- Dragos, K., & Smarsly, K. (2015). *Embedding numerical models into wireless sensor nodes for structural health monitoring*. Paper presented at the Proceedings of the 10th International Workshop on Structural Health Monitoring (IWSHM), Stanford, CA, USA.
- Duvenaud, D. (2014). *Automatic model construction with Gaussian processes*. University of Cambridge,
- Ering, P., & Babu, G. L. S. (2016). Probabilistic back analysis of rainfall induced landslide- A case study of Malin landslide, India. *Engineering Geology*, *208*, 154-164. doi:10.1016/j.enggeo.2016.05.002
- Fabbrocino, G., Laorenza, C., Rainieri, C., & Santucci de Magistris, F. (2009). *Seismic monitoring of structural and geotechnical integrated systems*. Paper presented at the Mater Forum.
- Feroz, F., & Hobson, M. P. (2008). Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *Monthly Notices of the Royal Astronomical Society*, *384(2)*, 449-463. doi:10.1111/j.1365-2966.2007.12353.x
- Feroz, F., Hobson, M. P., & Bridges, M. (2009). MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society*, *398(4)*, 1601-1614. doi:10.1111/j.1365-2966.2009.14548.x
- Forrester, A., Sobester, A., & Keane, A. (2008). *Engineering design via surrogate modelling: a practical guide*: John Wiley & Sons.
- Gens, A., Ledesma, A., & Alonso, E. E. (1996). Estimation of parameters in geotechnical backanalysis — II. Application to a tunnel excavation problem. *Computers and Geotechnics*, *18(1)*, 29-46. doi:[https://doi.org/10.1016/0266-352X\(95\)00022-3](https://doi.org/10.1016/0266-352X(95)00022-3)

- Hohenbichler, M., & Rackwitz, R. J. S. s. (1982). First-order concepts in system reliability. *1*(3), 177-188.
- Honfi, D. (2013). *Design for Serviceability-A probabilistic approach*: Lund University.
- Honjo, Y., Wen-Tsung, L., & Guha, S. (1994). Inverse analysis of an embankment on soft clay by extended Bayesian method. *International Journal for Numerical and Analytical Methods in Geomechanics*, *18*(10), 709-734. doi:10.1002/nag.1610181004
- Hoyle, N. (2006). *Automated multi-stage geometry parameterization of internal fluid flow applications*. University of Southampton,
- Ivanov, L. D. (2013). Semi-probabilistic approach for assessment of the requirement for minimum hull girder section modulus. *Ships and Offshore Structures*, *8*(1), 45-54. doi:10.1080/17445302.2011.605691
- JCSS. (2000). JCSS Probabilistic Model Code. Part III - Resistance Models. Retrieved from http://www.jcss.byg.dtu.dk/-/media/Subsites/jcss/english/publications/probabilistic_model_code/part_iii.ashx?la=da
- JCSS. (2001). JCSS Probabilistic model code.
- Jonkman, S., Steenbergen, R., Morales-Napoles, O., Vrouwenvelder, A., & Vrijling, J. J. C. C. (2015). Probabilistic Design: Risk and Reliability Analysis in Civil Engineering.
- Kandasamy, K., Schneider, J., & Póczos, B. (2017). Query efficient posterior estimation in scientific experiments via Bayesian active learning. *Artificial Intelligence*, *243*, 45-56. doi:10.1016/j.artint.2016.11.002
- Kelly, R., & Huang, J. S. (2015). Bayesian updating for one-dimensional consolidation measurements. *Canadian Geotechnical Journal*, *52*(9), 1318-1330. doi:10.1139/cgj-2014-0338
- Lang, B., Poppe, T., Minin, A., Mokhov, I., Kuperin, Y., Mekler, A., & Liapakina, I. (2008). Neural clouds for monitoring of complex systems. *Optical Memory and Neural Networks*, *17*(3), 183-192. doi:10.3103/S1060992X08030016
- Lassing, B., & Barneveld, A. <https://waterenklimaat.nl/en/research-tracks/future-proof-hydraulic-structures/>.
- Ledesma, A., Gens, A., & Alonso, E. E. (1996). Estimation of parameters in geotechnical backanalysis — I. Maximum likelihood approach. *Computers and Geotechnics*, *18*(1), 1-27. doi:[https://doi.org/10.1016/0266-352X\(95\)00021-2](https://doi.org/10.1016/0266-352X(95)00021-2)
- Levasseur, S., Malecot, Y., Boulon, M., & Flavigny, E. (2008). Soil parameter identification using a genetic algorithm. *International Journal for Numerical and Analytical Methods in Geomechanics*, *32*(2), 189-213. doi:10.1002/nag.614

- Levasseur, S., Malécot, Y., Boulon, M., & Flavigny, E. (2007). Soil parameter identification using a genetic algorithm. *International Journal for Numerical and Analytical Methods in Geomechanics*, 32(2), 189-213. doi:10.1002/nag.614
- Li, J., & Marzouk, Y. M. (2014). Adaptive Construction of Surrogates for the Bayesian Solution of Inverse Problems. *SIAM Journal on Scientific Computing*, 36(3), A1163-A1186. doi:10.1137/130938189
- Li, X. Y., Zhang, L. M., & Jiang, S. H. (2016). Updating performance of high rock slopes by combining incremental time-series monitoring data and three-dimensional numerical analysis. *International Journal of Rock Mechanics and Mining Sciences*, 83, 252-261. doi:10.1016/j.ijrmms.2014.09.011
- MacKay, & JC, D. (2003). *Information theory, inference and learning algorithms*: Cambridge university press.
- Meo, M., & Zumpano, G. (2005). On the optimal sensor placement techniques for a bridge structure. *Engineering Structures*, 27(10), 1488-1497. doi:10.1016/j.engstruct.2005.03.015
- Miro, S., Konig, M., Hartmann, D., & Schanz, T. (2015). A probabilistic analysis of subsoil parameters uncertainty impacts on tunnel-induced ground movements with a back-analysis study. *Computers and Geotechnics*, 68, 38-53. doi:10.1016/j.compgeo.2015.03.012
- Moreira, N., Miranda, T., Pinheiro, M., Fernandes, P., Dias, D., Costa, L., & Sena-Cruz, J. (2013). Back analysis of geomechanical parameters in underground works using an Evolution Strategy algorithm. *Tunnelling and Underground Space Technology*, 33, 143-158. doi:10.1016/j.tust.2012.08.011
- Mukherjee, P., Parkinson, D., & Liddle, A. R. (2006). A nested sampling algorithm for cosmological model selection. *Astrophysical Journal*, 638(2), L51-L54. doi:Doi 10.1086/501068
- Naves, T. (2018). POVM Eemdijkproef: Factual report Opbouw en Proef POT (Product Q).
- Neal, R. M. (2011). MCMC Using Hamiltonian Dynamics. *Handbook of Markov Chain Monte Carlo*, 2(11), 113-162. doi:BOOK_Doi 10.1201/b10905
- Owen, N. E., Challenor, P., Menon, P. P., & Bennani, S. (2017). Comparison of Surrogate-Based Uncertainty Quantification Methods for Computationally Expensive Simulators. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1), 403-435. doi:10.1137/15m1046812

- Papadimitriou, C. (2004). Optimal sensor placement methodology for parametric identification of structural systems. *Journal of Sound and Vibration*, 278(4-5), 923-947. doi:10.1016/j.jsv.2003.10.063
- Papadimitriou, C., Beck, J. L., & Au, S. K. (2000). Entropy-based optimal sensor location for structural model updating. *Journal of Vibration and Control*, 6(5), 781-800. doi:10.1177/107754630000600508
- Papadimitriou, C., & Lombaert, G. (2012). The effect of prediction error correlation on optimal sensor placement in structural dynamics. *Mechanical Systems and Signal Processing*, 28, 105-127. doi:10.1016/j.ymsp.2011.05.019
- Park, D., & Park, E. (2017). *Application of Response Surface Methodology to Geotechnical Parameter Estimation in Tunneling*. Paper presented at the 51st US Rock Mechanics/Geomechanics Symposium.
- Park, H. S., Lee, H. M., Adeli, H., & Lee, I. (2007). A new approach for health monitoring of structures: Terrestrial laser scanning. *Computer-Aided Civil and Infrastructure Engineering*, 22(1), 19-30. doi:10.1111/j.1467-8667.2006.00466.x
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(Oct), 2825-2830.
- Peeters, B. (2000). System identification and damage detection in civil engineering.
- Phoon, K.-K., & Tang, C. (2019). Characterisation of geotechnical model uncertainty. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 13(2), 101-130. doi:10.1080/17499518.2019.1585545
- Rackwitz, R. (2000). Reviewing probabilistic soils modelling. *Computers and Geotechnics*, 26(3-4), 199-223. doi:10.1016/S0266-352x(99)00039-7
- Razavi, S., Tolson, B. A., & Burn, D. H. J. W. R. R. (2012). Review of surrogate modeling in water resources. *48*(7).
- Rijkswaterstaat. The flood of 1953. Retrieved from <https://www.rijkswaterstaat.nl/english/water/water-safety/the-flood-of-1953/index.aspx>
- SAAF. (2019). SAAF manual. Retrieved from <http://saaf.measurand.com/home>
- Schweckendiek, T. (2006). Structural reliability applied to deep excavations. Coupling reliability methods with finite elements.
- Schweckendiek, T., & Kanning, W. (2016). *Reliability Updating for Slope Stability of Dikes: Approach with Fragility Curves (background Report)*: Deltares.

- Shaw, J. R., Bridges, M., & Hobson, M. P. (2007). Efficient Bayesian inference for multimodal problems in cosmology. *Monthly Notices of the Royal Astronomical Society*, 378(4), 1365-1370. doi:10.1111/j.1365-2966.2007.11871.x
- Simon, J. L. (1997). *Resampling: The New Statistics* (2nd ed.).
- Skilling, J. (2006). Nested Sampling for General Bayesian Computation. *Bayesian Analysis*, 1(4), 833-859. doi:Doi 10.1214/06-Ba127
- Slomp, R. J. W. (2012). Flood risk and water management in the Netherlands.
- Sykora, M., Holicky, M., & Markova, J. (2013). Verification of existing reinforced concrete bridges using the semi-probabilistic approach. *Engineering Structures*, 56, 1419-1426. doi:10.1016/j.engstruct.2013.07.015
- Takhtaganov, T., & Müller, J. J. a. p. a. (2018). Adaptive Gaussian process surrogates for Bayesian inference.
- Tang, Y. G., & Kung, G. T. C. (2009). Application of nonlinear optimization technique to back analyses of deep excavation. *Computers and Geotechnics*, 36(1-2), 276-290. doi:10.1016/j.compgeo.2008.02.004
- Teixeira, A., Rippi, K., Schweckendiek, T., Brinkman, H., Nuttall, J., Hellebrandt, L., & Courage, W. J. R. P. N. K. v. d. T. (2015). Soil-structure interaction–reliability analysis of a retaining wall.
- Union, E. (2009). The economics of climate change adaptation in EU coastal areas. Retrieved from https://ec.europa.eu/maritimeaffairs/sites/maritimeaffairs/files/docs/body/executive_summary_en.pdf
- Verruijt, A. (2012). *Soil Mechanics*: Delft University of Technology.
- Wang, H., & Li, J. (2018). Adaptive Gaussian Process Approximation for Bayesian Inference with Expensive Likelihood Functions. *Neural Comput*, 30(11), 1-23. doi:10.1162/neco_a_01127
- Wang, L., Hwang, J. H., Luo, Z., Juang, C. H., & Xiao, J. (2013). Probabilistic back analysis of slope failure – A case study in Taiwan. *Computers and Geotechnics*, 51, 12-23. doi:<https://doi.org/10.1016/j.compgeo.2013.01.008>
- Wang, L., Hwang, J. H., Luo, Z., Juang, C. H., Xiao, J. J. C., & Geotechnics. (2013). Probabilistic back analysis of slope failure—a case study in Taiwan. 51, 12-23.
- Williams, C. K., & Rasmussen, C. E. (2006). *Gaussian processes for machine learning* (Vol. 2): MIT Press Cambridge, MA.
- Wolters, H., Bakker, K., & De Gijt, J. (2013). *Reliability of quay walls using finite element analysis calibration of partial safety factors in quay wall design by probabilistic plaxis*

- calculations*. Paper presented at the EACEF 2013: Proceedings of the 4th International Conference of the Euro-Asia Civil Engineering Forum, Singapore, 26-28 June 2013.
- Worden, K., & Manson, G. (2007). The application of machine learning to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851), 515-537. doi:10.1098/rsta.2006.1938
- Xu, B., & Low, B. K. (2006). Probabilistic stability analyses of embankments based on finite-element method. *Journal of Geotechnical and Geoenvironmental Engineering*, 132(11), 1444-1454. doi:10.1061/(ASCE)1090-0241(2006)132:11(1444)
- Yi, T. H., Li, H. N., & Gu, M. (2011). Optimal sensor placement for structural health monitoring based on multiple optimization strategies. *Structural Design of Tall and Special Buildings*, 20(7), 881-900. doi:10.1002/tal.712
- Zhang, J., Tang, W. H., Zhang, L. J. J. o. G., & Engineering, G. (2009). Efficient probabilistic back-analysis of slope stability model parameters. *136*(1), 99-109.
- Zhang, J., Tang Wilson, H., & Zhang, L. M. (2010). Efficient Probabilistic Back-Analysis of Slope Stability Model Parameters. *Journal of Geotechnical and Geoenvironmental Engineering*, 136(1), 99-109. doi:10.1061/(ASCE)GT.1943-5606.0000205
- Zhang, L., Wu, F., Zheng, Y., Chen, L., Zhang, J., & Li, X. (2018). Probabilistic calibration of a coupled hydro-mechanical slope stability model with integration of multiple observations. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 12(3), 169-182. doi:10.1080/17499518.2018.1440317
- Zhang, L. L., Zhang, J., Zhang, L. M., & Tang, W. H. (2010). Back analysis of slope failure with Markov chain Monte Carlo simulation. *Computers and Geotechnics*, 37(7-8), 905-912. doi:10.1016/j.compgeo.2010.07.009

Annex A Used Python packages

This annex summarizes the most important Python packages used for the analyses documented in this thesis.

	Package name	Comments
Plaxis control	plxscripting	Related to API of Plaxis.
Surrogate modelling	scikit-learn	Popular and free machine learning package.
	SMT	Consisting of different surrogate modeling methods, e.g. Kriging, least square approximation, etc.
Bayesian inference	pymc3	Used for Bayesian statistical modeling and probabilistic machine learning including advanced MCMC and variational inference algorithms.
	nestle	Implementation of nested sampling (also used for reliability assessment).
	dynesty	Implementation of nested sampling, including evidence-oriented static nested sampling and posterior-oriented dynamic nested sampling (also can be used for reliability assessment).
Reliability assessment	openturns	Dealing with uncertainty propagation by probabilistic methods, including data analysis, probabilistic modelling, reliability, etc.
	pyre	Used for structural reliability analysis.

Annex B Sheet pile wall structures

Sheet pile walls are retaining walls that are designed and used to resist the lateral pressure and/or potential lateral deformation of adjacent materials, e.g. soil and water. They are widely used in wharfs, dams and river bank protection. A sheet pile wall is “a vertical wall consisting of long thin elements (steel, concrete or wood), that are being driven into the ground. The elements are usually connected by joints, consisting of special forms of the element at the two ends. Compared to a massive wall (of concrete or stone), a sheet pile wall is a flexible structure, in which bending moments will be developed by the lateral load, and that should be designed so that they can withstand the largest bending moments” (Verruijt, 2012). Anchors are often used to strengthen sheet pile walls. In Figure 0.1 there are some typical sheet pile walls while in Figure 0.2 different type of anchor layouts are included.

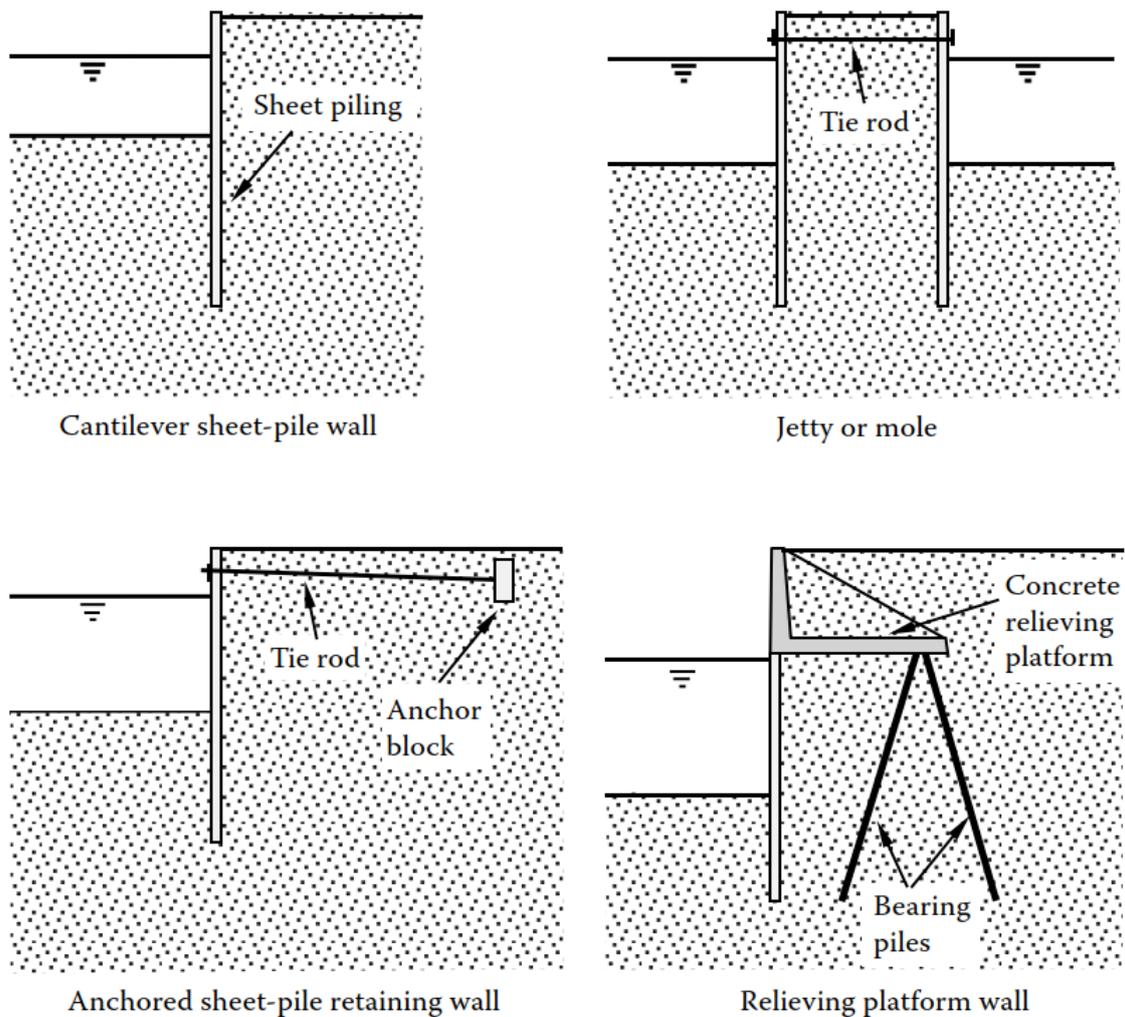


Figure 0.1: Cross-sections through some typical permanent sheet-pile structures (Clayton, Woods, Bond, & Milititsky, 2014).

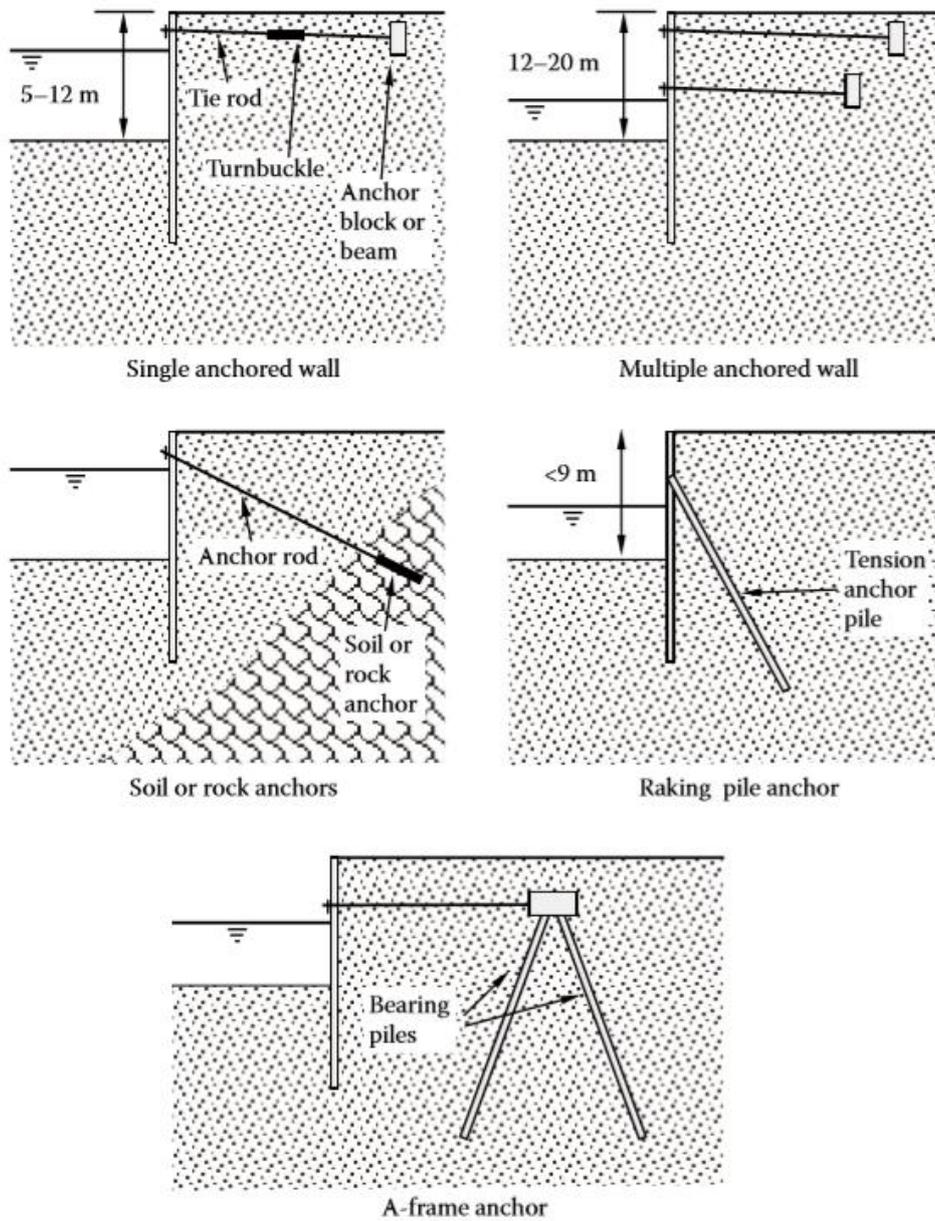


Figure 0.2: Anchored sheet pile wall schemes (Clayton et al., 2014).

Many materials can be used in the construction of sheet pile walls, of which the three most commonly used are steel, wood and concrete and their main features are summarized in Table 0.1 based on (Clayton et al., 2014).

Table 0.1: Main features of different sheet pile wall materials.

Used materials	Main features
Steel	<p>Variety of cross sections with a wide range of strength;</p> <p>Economy;</p> <p>Lack of buckling under heavy driving;</p> <p>Availability in different combinations to increase wall section modulus;</p> <p>Reusability for temporary works;</p> <p>Relatively light weight;</p> <p>The possibility of increasing the pile length by welding or bolting;</p> <p>Suffering from corrosion.</p>
Wood	<p>Usually for temporary work;</p> <p>Short spans;</p> <p>Short life expectancy.</p>
Concrete	<p>Variety of cross sections;</p> <p>Durable;</p> <p>Thick section required;</p> <p>Large soil displacement and driving resistance during construction.</p>

Sheet pile walls are prevalent structures in the Dutch water defense system, since a considerable area of the Netherlands is below sea level and a large part of its population lives there. The existing sheet pile wall are aging and require a regular inspection and assessment to ensure that the water defense system meets a prescribed safety level. This consideration motivates this work.