University of London
Imperial College of Science, Technology and Medicine
Department of Civil and Environmental Engineering

Data-driven uncertainty quantification on offshore piles

Ningxin Yang

Submitted on December 14, 2023

Abstract

Parameter identification is crucial for ensuring the permanence and stability of offshore structures. However, the uncertainties associated with parameters in the field significantly impact the behaviors of the piles. Uncertainties in the geotechnical context arise from several factors, such as differences between in-situ tests and laboratory experiments, spatial variations in the soil profile, and the rationality of the constitutive model. Additionally, uncertainties are also linked to changes in dimensions during the construction phase and variations in the magnitude of design loads. Manual back analysis or traditional design method based on Monte Carlo, are time-consuming and labor-intensive. Although the Bayesian framework offers insights into understanding and updating uncertainties, the computational demands of inference analysis pose substantial challenges for predictive purposes.

Nowadays, data-driven methodologies have permeated various engineering domains, facilitating automatic evolution over time to mitigate uncertainties. However, in the realm of geotechnical engineering, there exists a notable absence of guidelines for data-driven uncertainty quantification (UQ). Starting from a pile issue, this thesis endeavors to address this gap by delving into the latest research and hopes to put forth a comprehensive UQ design framework that can be generally applied in geotechnical engineering.

Contents

List of Tables

List of Figures

Nomenclature

$\boldsymbol{x}$	Vector of input parameter
$\boldsymbol{\varepsilon}$	Gaussian discrepancy
$\boldsymbol{x}^{\mathrm{MAP}}$	Maximum a posterior estimate
$\boldsymbol{x}^{\mathrm{mean}}$	posterior mean
$\boldsymbol{x}^{\mathrm{ML}}$	Maximum likelihood estimate
$\boldsymbol{x}^{s}_{t\_N_{chain}}$	s_{th} samples from N_{th} chain at t_{th} stage
$\boldsymbol{x}_t^{(\star)}$	Candidate samples at t_{th} stage
$\boldsymbol{x}_t^s$	s_{th} samples from input space at t_{th} stage
$\boldsymbol{y}$	Vector of particular evaluation/observation
$\hat{S}_{eff}$	Effective sample size
$\mathcal{D}_{\boldsymbol{X}}$	Input parameters space
$\mathcal{L}(\boldsymbol{x};\mathcal{Y})$	Likelihood function
$\mathcal{M}$	Computational model
$\mathcal{R}$	Residual
$\mathcal{X}_{t\_N_{chain}}$	Samples from N_{th} chain at t_{th} stage
$\mathcal{X}_t$	Samples at t_{th} stage
$\mathcal{Y}$	Set of evaluations/observations across all time stages

$\mathcal{Y}_{\mathrm{t}}$ Evaluation/observation at t_{th} stage

$\pi(\boldsymbol{x})$ Prior

$\pi(\cdot \mid \cdot)$ Conditional Probability density

$\pi(\cdot)$ Probability density

$\pi(\mathcal{Y})$ Evidence

$\tilde{\mathcal{M}}$ Surrogate model

$q(\boldsymbol{x})$ Proposal distribution

S_{min} Threshold effective sample size

Acronyms

AIES Affine invariant ensemble sampler. 19–21, 25

CDF Cumulative Density Function. 11, 14

CI Confidence Interval. 11

COV Coefficient of variance. ix, 11

DR Dimensionality reduction. ix, 29, 35–37

MAP Maximum a Posterior. 9, 10

MCMC Markov Chain Monte Carlo. 13, 17, 19–21, 24–26, 29

MH Metropolish-Hasting. ix, 17–19

ML Maximum Likelihood. 10

PCA Principal component analysis. 35, 36

PCE Polynomial chaos expansion. vi, 30, 31, 33–36, 39

PDF Probability Density Function. 7, 9, 11, 17, 20, 31

PGM Probabilistic graphical model. 40

POPGM Partially observed probabilistic graphical model. ix, 40–42

QoI Quantities of interest. 27

SIS Sequential importance sampling. ix, 23–25

SISR Sequential importance sampling and resampling. ix, 24, 25

SMC Sequential Monte Carlo. 13, 21, 22, 24, 25

Chapter 1

Introduction

1.1 Background

Offshore monopiles are increasingly preferred for wind farm installations, owing to their benefits in clean energy generation and convenient deployment. These cylindrical steel structures are driven into the seabed to establish a stable foundation for wind turbines. While offshore wind energy holds promise as a source of clean and sustainable power, the adoption of offshore monopiles introduces specific engineering challenges. A primary concern associated with offshore monopiles is addressing potential issues related to excessive pile displacements induced during both their installation and operation phases (Byrne & Houlsby, 2003; Randolph et al., 2005). Excessive movements in the piles can result in significant displacements and rotations within supporting structures, subsequently leading to damage or structural instability. Consequently, accurate prediction and effective management of pile deformations become paramount when designing and analyzing support systems for offshore monopile installations.

After soil properties, design loads and pile dimensions are acquired from a technical report, estimating the pile response is typically done through empirical solutions in guidelines or numerical simulations. Several design methods have been provided in the design codes (API, 2011; Bhattacharya, 2019) to predict offshore pile $p - y$ curve. However, it is challenging to incorporate all influential factors, such as pile length, soil layer, soil

properties and loading conditions into a simplified empirical model. Rapid advancement of computational techniques has facilitated the numerical models (Randolph & Gourvenec, 2017; Taborda et al., 2020; Zdravković et al., 2020; Royston et al., 2022), serving as a potent analytical tool. Nonetheless, it necessitates a substantial number of simulation runs. This presents a significant challenge on the back-calculation of parameter properties, especially in the context of inverse and predictive analysis.

In recent research endeavors, Bayesian probability frameworks have garnered increasing recognition as an efficacious approach for inverse parameter estimation and response prediction (Finno & Calvello, 2005; Nakamura et al., 2011; Hsein Juang et al., 2013; Nguyen & Nestorović, 2016; Wagner et al., 2020; Jin et al., 2021; Tao et al., 2021; Buckley et al., 2023; Tang et al., 2023). In such circumstances, the Bayesian framework emerges as a powerful tool within the probabilistic context, facilitating parameter learning and informed decision-making. However, in high dimensions (typically in geotechnical areas), the computational cost of such model is high. Traditional Monte Carlo methods for UQ or manually overcoming this limitation is nearly impossible. There is an urgent requirement to tackle these issues. Luckily, the components necessary to address such problems have been already at our disposal. The UQ and machine learning communities have developed methodologies to efficiently address large-scale data models and disseminate uncertainties. However, in geotechnical engineering, there is a noticeable absense of guidlines for UQ area. Based on this, a data-driven dynamic UQ framework is hoped to be constructed and enable the real time information sharing between the models and measurements.

In practice, through adaptive Bayesian updating on the identified parameters and reduce the uncertainties on offshore piles, a field engineer would benefit from: (1) properly accounting the uncertainties of input variables; (2) real-time monitoring and adaptively predicting the pile response in probabilistic setting; (3) providing an efficient tool for data-driven decision making on pile operation and design. In the context of offshore engineering, in particular, it will show substantial potentials in various domains, including health monitoring, pile penetration and long-term bearing capacities (Wang et al., 2021;

Zhao et al., 2023; Stuyts et al., 2023).

1.2 Problem statement

Modern pile installation and proper estimation is becoming increasingly complex and vital to the reliability and permanence of the foundation in question. However, in the construction process, uncertainties and insufficient information (e.g., soil parameters) lead to inaccurate predictions of pile responses. The source of uncertainties may come from various reasons. Dealing with different uncertainties sources is a challenging task. One typical uncertainty source can be illustrated in Figure 1.1, which shows uncertainties sources:

- Fluctuating curve indicates the spatial variability
- Non-uniformity exists between in-situ test and laboratory experiment

Furthermore, geotechnical engineering problems inherently belong to the high-dimensional realm with substantial uncertainties. The substantial quantity of unknown distribution parameters might be too extensive for precise inference, given the small size of the accessible data, resulting in an underdetermined problem. Although some well-established approaches for fitting pile deformations are proposed to infer the underlying soil parameters and reduce the uncertainties, addressing this challenge becomes intricate when confronted with a multitude of input variables (i.e., $\mathcal{O}(10^2 - 10^4)$) (Lataniotis, 2019). Even if an adequate probabilistic input model can be obtained, performing the inference analysis through Monte Carlo simulation is still expensive. This poses challenges in understanding uncertainties and providing timely predictions for pile design. In these scenarios, it becomes practical to replace the computational model with a surrogate. Nevertheless, as dimensions increase, the efficacy of surrogate models diminishes, accompanied by escalating computational and storage expenses. This quandary is widely acknowledged as the *curse of dimensionality* (Verleysen & François, 2005). Computing the surrogate may become challenging, especially when confronted with a substantial number of input parameters (Lataniotis, 2019).

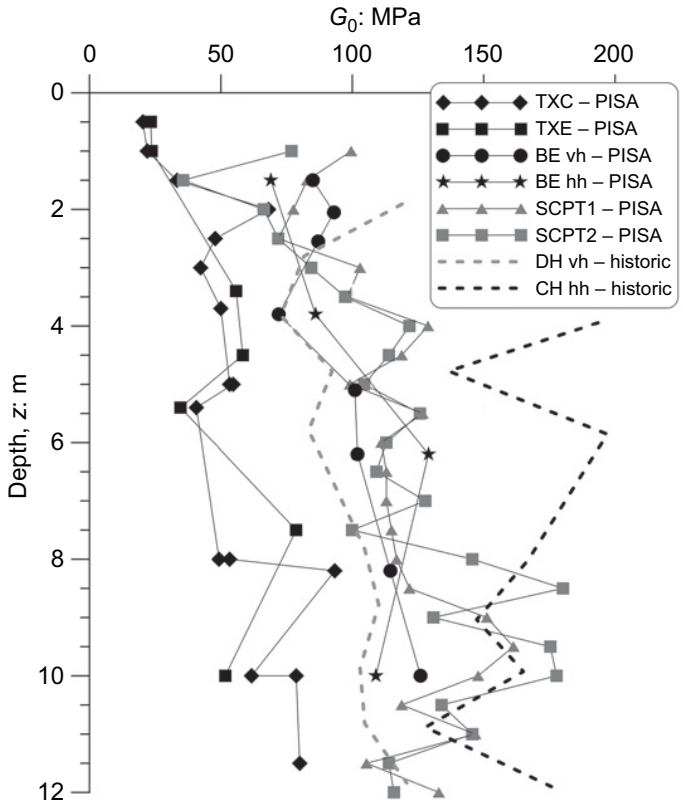


Figure 1.1: Stiffness characteristics at Cowden from Zdravković et al. (2020)

1.3 Objectives and outlines

Objectives:

This thesis aims to construct a resilient and scalable framework for uncertainty quantification, extending its applicability beyond offshore piles. This endeavor will leverage cutting-edge methodologies in surrogate modeling, uncertainty quantification, probabilistic graphical model and control theory, all geared towards facilitating extensive predictive digital twin capabilities. In particular, the specific goals of this thesis are:

- Develop a surrogate model suitable, not limited for offshore piles, for structures characterized by high input dimensions.
- Accelerate Bayesian inversion calculations for identified parameters to reduce the uncertainties, and providing real-time pile response predictions through adaptive enrichment of observed monitoring data.
- Develop an adaptive uncertainty quantification framework (not limited) for offshore

piles in a probabilistic way.

Outlines:

Chapter 2 introduces the fundamentals of Bayesian probabilistic theory. Chapter 3 discusses some of the most important forward and inverse UQ tools. Chapter 4 states the geotechnical UQ problems in our thesis and work plans for the next stage.

Chapter 2

Bayesian probabilistic theory

In probabilistic theory, two main interpretations prevail: frequentist and Bayesian. The frequentist perspective views probabilities as the long-term frequencies observed in infinite trials. On the other hand, the Bayesian perspective links probability to uncertainty and information, diverging from repeated trials.

Based on the quantity of accessible data, spanning from none to an infinite amount, diverse techniques can be employed:

- in cases where there is a lack of data to characterize the input parameters, a probabilistic model can be formulated solely based on expert judgment;
- when dealing with a substantial volume of data, one can fully leverage the tools of statistical inference, such as employing the method of moments (Wagner et al., 2020);
- when both expert judgment and very limited observations are available, Bayesian inference may be resorted to.

A notable benefit of the Bayesian interpretation lies in its ability to model events that lack long-term frequencies. Take, for example, the assessment of the probability of structural damage to a high-rise building, the collapse of a tunnel, or the occurrence of irreversible deformation in bridge piers. This event is anticipated to occur only a limited number of times over the structure's lifetime and is not expected to happen

repeatedly. Subsequently, it is important to quantify our uncertainty regarding this event and implement appropriate measures (see chapter 3 and chapter 4).

Since data collection is inherently constrained during the progression of most engineering projects (e.g., geotechnical problems), Bayesian theory stands out as a highly effective method. Therefore, this thesis exclusively explores Bayesian methods next, while detailed information on frequentist approaches can be found in Murphy (2012).

2.1 Bayesian inference

When dealing with a limited number of data points, direct statistical estimation becomes unreliable due to substantial statistical uncertainty in the sample estimates. In this context, Bayesian inference provides a solution by integrating prior knowledge on parameters with a small set of observed data points. Operating in this fully probabilistic setting, all unknowns are treated as random vectors. Distribution parameters can be denoted by $\boldsymbol{x}$ as realisations of the random vector $\boldsymbol{X} : \Omega \to \mathcal{D}_{\boldsymbol{X}}$. *Quantities of interest* gathered from output are gathered in a vector $\boldsymbol{y} \in \mathbb{R}^{N_{\text{out}}}$. The joint probability distribution of the combined random vector $(\boldsymbol{X}, \boldsymbol{Y}) : \Omega \to \mathcal{D}_{\boldsymbol{X}} \times D_{\boldsymbol{Y}}$ is represented by $\pi(\boldsymbol{x}; \boldsymbol{y})$. Leveraging the fundamental *sum rule* and *product rule* in probabilistic theory, the Probability Density Function (PDF) of the parameters $\boldsymbol{x}$ and the data $\boldsymbol{y}$ can be expressed as

$$\pi(\boldsymbol{x}|\boldsymbol{y}) = \frac{\mathcal{L}(\boldsymbol{x}; \boldsymbol{y}) \cdot \pi(\boldsymbol{x})}{\pi(\boldsymbol{y})} \tag{2.1}$$

which is also known as *Bayes' theorem* or *Bayes' rule*. In Bayesian terminology, this distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$ is called the posterior distribution and it is calculated by prior $\pi(\boldsymbol{x})$, likelihood $\mathcal{L}(\boldsymbol{x}; \boldsymbol{y}) \overset{\text{def}}{=} \pi(\boldsymbol{y}|\boldsymbol{x})$ and the evidence $\pi(\boldsymbol{y})$. These definitions of the likelihood function and evidence strictly hold only for a single data point $\mathcal{Y} = \{\boldsymbol{y}\}$, but can be generalised to multiple data points easily $\mathcal{Y} \overset{\text{def}}{=} \{\boldsymbol{y}^{(1)}, \cdots, \boldsymbol{y}^{(N)}\}$. These terms in Equation (2.1) have practical significance that we will briefly summarise next.

- Prior $\pi(\boldsymbol{x})$: In the Bayesian paradigm, before considering the data, the parameters $\boldsymbol{x}$ are treated as realisations derived from a random vector $\boldsymbol{X}$ presumed to adhere to the so-called prior distribution.

- Likelihood function $\mathcal{L}(\boldsymbol{x};\mathcal{Y})$: The likelihood function gauges the adequacy of the specified parametric distribution $\pi(\mathcal{Y}|\boldsymbol{x})$ describes the data. In most engineering cases, input parameters $\boldsymbol{x}$ are not measurable directly. To evaluate the likelihood $\mathcal{L}(\boldsymbol{x};\mathcal{Y})$, some ingredients are needed: a forward model $\mathcal{M}$, inferred input parameters $\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{X}}$, and a collection of experimental data $\mathcal{Y}$. The forward model $\boldsymbol{x} \rightarrow \boldsymbol{\mathcal{M}}(\boldsymbol{x})$ is a representation of a physical system under consideration. Thus, to establish a connection between the observations $\mathcal{Y}$ and model predictions, we introduce a *discrepancy term* $\boldsymbol{\varepsilon}$ and consider the following well-established format as:

$$\boldsymbol{y} = \mathcal{M}(\boldsymbol{x}) + \boldsymbol{\varepsilon} \tag{2.2}$$

where $\boldsymbol{\varepsilon} \in \mathbb{R}^{N_{\text{out}}}$ denotes the difference between the observation $\boldsymbol{y}$ and the prediction $\mathcal{M}(\boldsymbol{x})$. For the sake of simplicity, we consider it as an additive *Gaussian discrepancy* with zero mean and a covariance matrix $\boldsymbol{\Sigma}$ in this introduction:

$$\boldsymbol{\varepsilon} \in \mathcal{N}(\varepsilon|\mathbf{0}, \boldsymbol{\Sigma}) \tag{2.3}$$

If there are N independent measurements $\boldsymbol{y_i}$ collected in the dataset $\mathcal{Y} \overset{\text{def}}{=} \{\boldsymbol{y}_{(1)}, \cdots, \boldsymbol{y}_{(N)}\}$, the likelihood can expressed as follows:

$$\begin{aligned} \mathcal{L}(\boldsymbol{x};\mathcal{Y}) &= \prod_{i=1}^{N} N(\boldsymbol{y_i}|\mathcal{M}(\boldsymbol{x}), \boldsymbol{\Sigma}) \\ &= \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^{N_{\text{out}}} \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}\left(\boldsymbol{y_i} - \mathcal{M}(\boldsymbol{x})\right)^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{y_i} - \mathcal{M}(\boldsymbol{x})\right)\right) \end{aligned} \tag{2.4}$$

It is noted that simple Gaussian discrepancy assumption is only one out of many possible models. In a more general setting, other distributions for the discrepancy are used as well (Wagner et al., 2022). Due to the widespread used of the additive Gaussian models in engineering disciplines, this thesis is limited to Gaussian type.

- Evidence $\pi(\mathcal{Y})$: In Bayesian inference, $\pi(\mathcal{Y})$ is commonly regarded as a normalization

factor, ensuring that the posterior PDF can be integrated to one:

$$\pi(\mathcal{Y}) \stackrel{\text{def}}{=} \int_{\mathcal{D}_{\boldsymbol{X}}} \mathcal{L}(\boldsymbol{x};\mathcal{Y})\pi(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \tag{2.5}$$

A schematic Bayesian inference in two dimensional space is displayed in Figure 2.1. The plots show the various elements of the Bayesian inference procedure in the parameter and data spaces. We can see that in the parameter space, with new experimental data comes in, the posterior is more concentrated than the prior distribution.

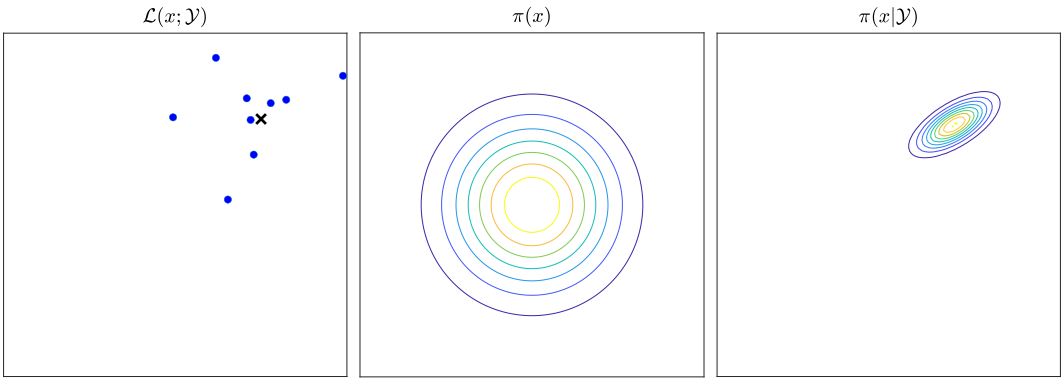


Figure 2.1: An Bayesian inference example in 2D space

2.2 Posterior quantities of interest

Under the Bayesian paradigm, the posterior distribution $\pi(\boldsymbol{x}|\mathcal{Y})$ is the solution to the inverse problem. However, the full distribution can contain too much information to allow statements about the inferred parameters. Therefore, it is common to process the posterior and extract certain quantities of interest that summarize the inversion results more concisely.

Point estimate

In some applications, one may be only interested in a single parameter, i.e., the one that characterise the inversion most suitably. Two most common *point estimation* methods are the *posterior mean* $\boldsymbol{x}^{\text{mean}}$ and *maximum a posterior* (MAP). The $\boldsymbol{x}^{\text{mean}}$ is given as:

$$\boldsymbol{x}^{\text{mean}} = \mathbb{E}[\boldsymbol{X}|\mathcal{Y}] = \int_{\mathcal{D}_{\boldsymbol{X}|\mathcal{Y}}} \boldsymbol{x}\pi(\boldsymbol{x}|\mathcal{Y})\mathrm{d}\boldsymbol{x} \tag{2.6}$$

It reflects what we expect the parameter mean value to be after the inference. The MAP parameter, on the other hand, is the one maximises the posterior:

$$\begin{aligned} \boldsymbol{x}^{\mathrm{MAP}} &= \arg\max_{\boldsymbol{x}\in\mathcal{D}_{\boldsymbol{X}}} \pi(\boldsymbol{x}|\mathcal{Y}) \\ &= \arg\max_{\boldsymbol{x}\in\mathcal{D}_{\boldsymbol{X}}} \mathcal{L}(\boldsymbol{x};\mathcal{Y})\pi(\boldsymbol{x}) \end{aligned} \tag{2.7}$$

where the evidence constant $\pi(\mathcal{Y})$ was omitted. The MAP point corresponds to the most likely value of the input parameters. It is closely related to the Maximum Likelihood (ML) point that is defined as

$$\boldsymbol{x}^{\mathrm{ML}} = \arg\max_{\boldsymbol{x}\in\mathcal{D}_{\boldsymbol{X}}} \mathcal{L}(\boldsymbol{x};\mathcal{Y}) \tag{2.8}$$

for which the forward model $\mathcal{M}$ produces the best agreement with the available data. Unlike the ML, the MAP point considers the prior information. The difference is typically larger in the case of little data, where the regularisation effect of the prior distribution is stronger. In case of uniform priors, the two are equal, provided that the ML point does not lie outside the prior support.

Posterior moments and *covariance*

Choices for the point estimation above disregards the estimation uncertainty in the parameters. Therefore, to more comprehensively characterise the posterior distribution and investigate the calibration, it is useful to compute the second *posterior moments* and the *covariance*. They are summarised in the *posterior covariance matrix* $\boldsymbol{C} \in \mathbb{R}^{M\times M}$ with entries:

$$\begin{aligned} \boldsymbol{C} &= \mathrm{Cov}[\boldsymbol{X}|\mathcal{Y}] \\ &= \int_{\mathcal{D}_{\boldsymbol{X}}} (\boldsymbol{x} - \mathbb{E}[\boldsymbol{X}|\mathcal{Y}])(\boldsymbol{x} - \mathbb{E}[\boldsymbol{X}|\mathcal{Y}])^{\mathsf{T}} \pi(\boldsymbol{x}|\mathcal{Y}) \mathrm{d}\boldsymbol{x} \end{aligned} \tag{2.9}$$

As the full posterior distribution is an M-dimensional object, which is inherently difficult to comprehend, one is typically also interested in the *posterior marginals* and *copula*

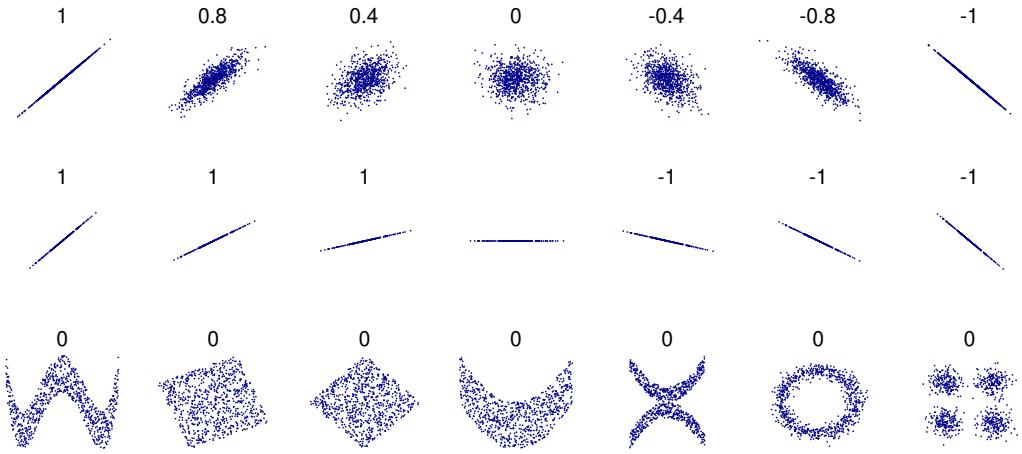


Figure 2.2: Example sets of (x, y) points with different COV from DenisBoigelot (2011)

dependence. The marginalised random variable $X_i|\mathcal{Y}$ is given by (see Equation (2.10)):

$$\pi(x_i|\mathcal{Y}) = \int_{\mathcal{D}_{\boldsymbol{X}_\mathbf{v}}} \pi(\boldsymbol{x}|\mathcal{Y}) \mathrm{d}\boldsymbol{x}_\mathbf{v}, \text{with } \mathbf{v} = \{1, \cdots, M\} \backslash i \tag{2.10}$$

This univariate PDF can be integrated to obtain the corresponding *cumulative density function* (CDF) which may then be used to define *confidence interval* (CI) on the calibrated parameters by means of quantiles. *Copula dependence* can be inferred easily from the joint PDF as $\boldsymbol{C}$. Different types of COV can be seen in Figure 2.2.

Predictive distribution

Next, to evaluate the predictive ability of the model, the Bayesian inference framework offers the possibility to compute *predictive distributions*. Using previously defined discrepancy model from Equation (2.3) and Equation (2.2), the *prior predictive* distribution can be represented as:

$$\pi(\boldsymbol{y}) = \int_{\mathcal{D}_{\boldsymbol{X}}} \pi(\boldsymbol{x}) \pi(\boldsymbol{y}|\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{2.11}$$

It summarises the uncertainty about the model output considering the discrepancy model before calibration. In practice, it is used to determine whether the measured data can be reproduced. It is essential for identifying and ruling out challenging inverse problems before proceeding with expensive calibration procedures. The *posterior predictive*

distribution (see Equation (2.12)) can be similarly written as:

$$\pi(\boldsymbol{y}|\mathcal{Y}) = \int_{\mathcal{D}_{\boldsymbol{X}|\mathcal{Y}}} \pi(\boldsymbol{x}|\mathcal{Y})\pi(\boldsymbol{y}|\boldsymbol{x})\mathrm{d}\boldsymbol{x} \tag{2.12}$$

2.3 Computational methods for Bayesian inference

Calculation for posterior distributions $\pi(\boldsymbol{x}|\mathcal{Y})$ is not easy. Because computing evidence $\pi(\mathcal{Y})$ is usually not a tractable problem, analytical solutions thus require more restrictions on the model. It can be only calculated analytically if it is given in a closed form. A common strategy we usually choose is a *conjugate prior* (Gelman et al., 1995) to the likelihood, so the integral can be represented analytically. For example, in a static Bayesian network, choices such as *variant elimination* and *belief propagation* (Murphy, 2012) can be seen. In the realm of a dynamic sequential model, *kalman filtering* (Nguyen & Nestorović, 2016) gives an closed form for the parameter identification. However, in the general cases, the solution to the posterior can be rarely analytical due to a complex model or high computation for the evidence $\pi(\mathcal{Y})$. Thus, we need to resort to approximation method.

There are usually two categories of approximation methods: *optimisation based approximation* and *Monte Carlo sampling* methods.

- *Optimisation based approximation*: This method usually refers to variational inference. The basic principle is adopting some analytical distributions to approximate the posterior based on some loss functions. Then we can measure the similarity (e.g., *Kullback-Leibler divergence*) between two distributions.

  The advantages of optimised approximation methods are: (1) computationally efficient and work well on large models; (2) has absolute converging criteria which makes easy to determine when to stop the modelling; (3) scales better and are more amenable to parallelization. However, there are some problems itself: (1) in contrast to methods based on sampling, variational approaches are unlikely to discover the globally optimal solution; (2) the precision of variational approaches is frequently constrained by the

structure of the approximation.

- *Monte Carlo sampling*: Sampling method is another way to approximate the posterior distribution. Examples include *inverse probability transform*, *rejection sampling*, *importance sampling*, *Markov chain Monte Carlo* (MCMC) and *sequential Monte Carlo* SMC. These techniques produce random samples from a *proposal distribution*, utilizing them to estimate both the posterior distribution and the associated statistics.

  Monte Carlo sampling has the advantages that: (1) it is more straightforward and flexible; (2) given sufficient time and samples, it is assured to identify the globally optimal solution. However, for a good accuracy, *Monte Carlo sampling* takes more time in the calculation and require sophisticate deployments on choosing an appropriate sampling technique.

In short mentioned above, *Monte Carlo sampling* is asymptotically exact, accurately approximating the target distribution with increasing samples. In contrast, *Optimisation based approximation* lacks guarantees but yields faster results. Both *optimisation based approximation* and *Monte Carlo sampling* are significant topics. Current studies on these are still very active with numerous techniques proposed in the past few years. More details can be found in Murphy (2012) and Blei et al. (2017).

It is worthy noted that it is not always the case *Monte Carlo sampling* is better than *optimisation based approximation* or vice versa. The choice for approximate a posterior is totally problem-specified. Faster posterior approximation requires trading off additional accuracy. In this thesis, however, we expect to find the global optimal values and hope the solution with guarantee. Therefore, our thesis is expected to focus only on *Monte Carlo sampling* as detailed in the next section.

2.4 Monte Carlo sampling

In this section, we will explore algorithms belonging to the category of *Monte Carlo approximation*. The idea is straightforward: generate s samples from the t_{th} step posterior, $\boldsymbol{x}_t^s \sim \pi(\boldsymbol{x}_t|\mathcal{Y}_t)$, in which t represents the time or stage in a dynamic model

(state-space model) and $\mathcal{Y}_t$ denotes the observation at t_{th} step. Because most engineering projects are carried out in stages as shown in Figure 2.3, a recursive updating and sampling scheme is required across the multiple stages in the dynamic model. The introduction of subscript t aims to elucidate the sampling process at different stages. Following that, these samples are employed to compute any quantity of interest $\mathbb{E}[f|\mathcal{Y}_t] \approx \frac{1}{S}\sum_{s=1}^{S} f(\boldsymbol{x}_t^s)$ through an appropriate function f. By getting a sufficient number of samples, we can attain the desired level of accuracy.

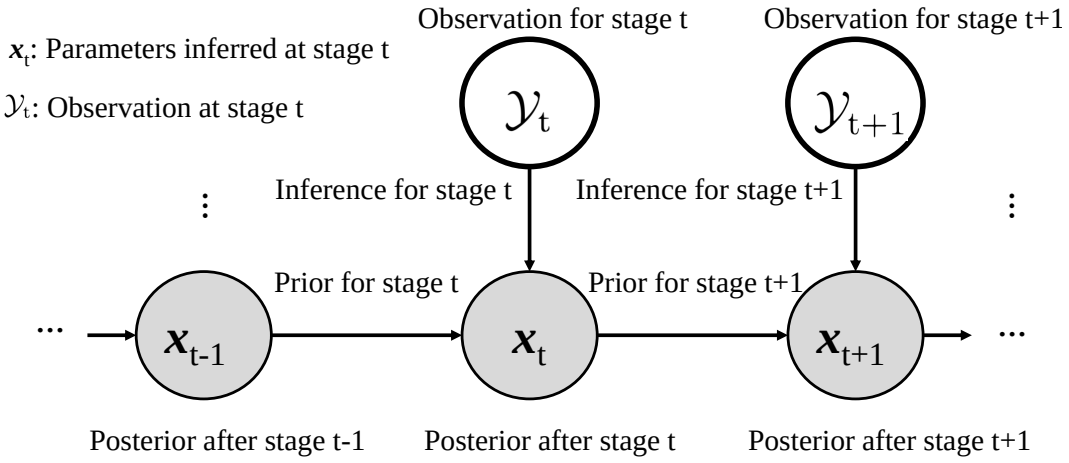


Figure 2.3: Sequential Bayesian updating

2.4.1 Inverse probability transform

The most straightforward technique for sampling can be *inverse probability transform* from a univariate distribution. As shown in Figure 2.4, if we can get the *cumulative density function* (CDF), then we can easily generate samples by computing $\boldsymbol{x}_t^s = \text{CDF}(\mathcal{U})$. $\mathcal{U}$ follows uniform distribution $\mathcal{U} \sim U(0,1)$ using a *pseudo random number generator*.

2.4.2 Rejection sampling

In cases where inverse CDF method is not feasible, a straightforward alternative is to employ *rejection sampling*. We propose a distribution $q(\boldsymbol{x})$ ensuring $Mq(\boldsymbol{x}) \geq \tilde{\pi}(\boldsymbol{x})$, where M is a constant. Here, $\tilde{\pi}(\boldsymbol{x})$ represents unnormalized $\pi(\boldsymbol{x})$ (denoted as $\pi(\boldsymbol{x}) = \tilde{\pi}(\boldsymbol{x})/Z$ for some unknown constant Z). The function $Mq(\boldsymbol{x})$ serves an upper bound for $\tilde{\pi}$. At time stage t, we sample $\boldsymbol{x}_t^s \sim q(\boldsymbol{x}_t)$, representing the random selection of a position $\boldsymbol{x}_t^s$. Subsequently, we sample $u \sim U(0,1)$, representing a random height under the upper

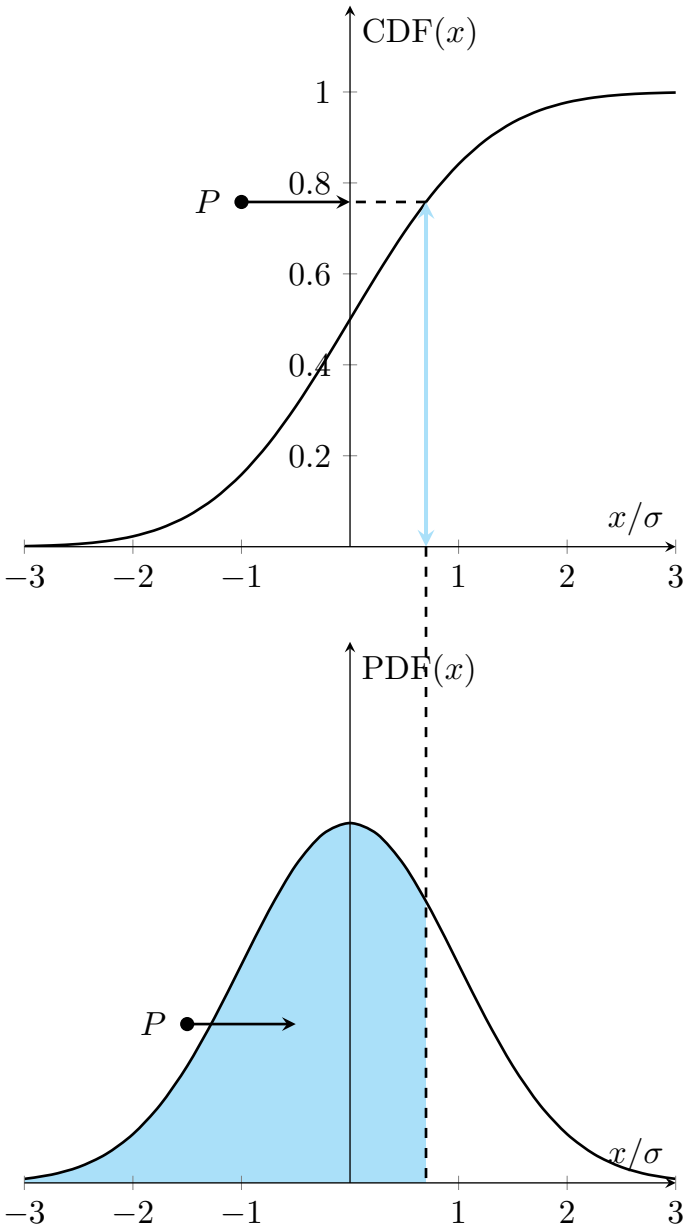


Figure 2.4: Sampling using an inverse CDF

bound. If $u \geq \frac{\tilde{\pi}(\boldsymbol{x}_t^s)}{Mq(\boldsymbol{x}_t^s)}$, the sample is rejected, otherwise it is accepted. The shaded region in Figure 2.5 shows the acceptance region, and the unshaded region indicates the rejection area. However, in high-dimensional spaces, the majority of the space tends to be empty. Consequently, the likelihood of this method accepting a point can be remarkably low when dealing with multidimensional spaces.

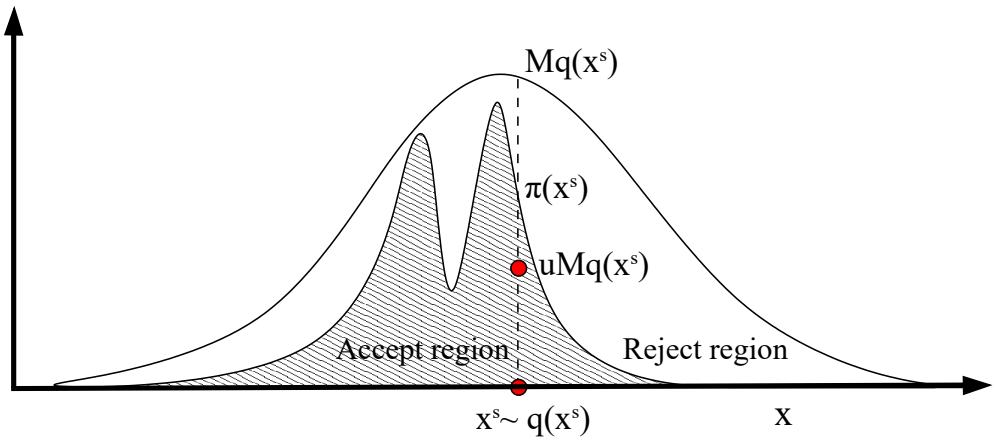


Figure 2.5: Schematic illustration of *rejection sampling* from Andrieu et al. (2003)

2.4.3 Importance sampling

Now, we introduce another Monte Carlo technique called *importance sampling*, designed for estimating integrals of the following form:

$$E[f(\boldsymbol{x}_t)] = \int f(\boldsymbol{x}_t)\pi(\boldsymbol{x}_t)d\boldsymbol{x}_t \approx \frac{1}{S}\sum_{s=1}^{S} f(\boldsymbol{x}_t^s) \tag{2.13}$$

where $\boldsymbol{x}_t \sim \pi(\boldsymbol{x}_t)$. As mentioned, the idea of *Monte Carlo sampling* is to simply sample $\boldsymbol{x}_t$ from the distribution $\pi(\boldsymbol{x}_t)$. However, when it comes to tricky targets $\pi(\boldsymbol{x}_t)$, the above estimation can be less efficient, meaning it requires huge amounts of samples to ensure the accuracy. In this case, it is more advantageous to sample from a proposed distribution $q(\boldsymbol{x}_t) \propto f(\boldsymbol{x}_t)\pi(\boldsymbol{x}_t)$ than directly from $\pi(\boldsymbol{x}_t)$.

The main concept behind the *importance sampling* is to draw samples at the t_{th} step from a proposed distribution $q(\boldsymbol{x}_t)$ and adjust the the integral using importance weights to the targets. It then uses these samples to estimate the integral as follows:

$$E[f(\boldsymbol{x}_t)] = \int f(\boldsymbol{x}_t)\pi(\boldsymbol{x}_t)d\boldsymbol{x}_t = \int f(\boldsymbol{x}_t)\frac{\pi(\boldsymbol{x}_t)}{q(\boldsymbol{x}_t)}q(\boldsymbol{x}_t)d\boldsymbol{x}_t \approx \frac{1}{S}\sum_{s=1}^{S} f(\boldsymbol{x}_t^s)\frac{\pi(\boldsymbol{x}_t^s)}{q(\boldsymbol{x}_t^s)} \tag{2.14}$$

where $\frac{\pi(\boldsymbol{x}_t^s)}{q(\boldsymbol{x}_t^s)}$ is the importance weight. With a proper proposal distribution $q(\boldsymbol{x}_t)$, the integral estimation can be close as possible to the true value (improving the approximation means decreasing the variance of estimation $\mathrm{Var}(\boldsymbol{x}_t) = \mathbb{E}[{\boldsymbol{x}_t}^2] - \mathbb{E}[\boldsymbol{x}_t]^2$). However, in

practice, the choice for the $q(\boldsymbol{x}_t^s)$ is not satisfying. This is a general difficult task, especially in high dimensions.

2.4.4 Markov Chain Monte Carlo

Closed-form expressions for the posterior distribution in Equation (2.1) cannot be derived in general. The exceptions are some linear models with conjugate distributions which were mentioned in section 2.3. Compared with non-iterative methods above, Markov Chain Monte Carlo (MCMC) are a category of algorithms enabling the sampling from the posterior by constructing a Markov chain. This process asymptotically converges to the true PDF.

MCMC methods were a breakthrough for solving Bayesian inverse problems and since then it has gained tremendous popularity in engineering areas. In a *SIAM News* survey, MCMC was listed as one of the 10 most important algorithms of the 20th century. Originally developed by statistical mechanics, it is used to construct *Markov chains* to sample a selected *target distribution*. Within the framework of Bayesian updating, this *target distribution* is the posterior in our analysis. Based on the *Markov chains*, every stage depends only on the last previously attained stage. We consider here one discrete *Markov chain* of the type $\mathcal{X}_t = \{\boldsymbol{x}_t^1, \cdots, \boldsymbol{x}_t^{N_\mathcal{X}}\}$, in which $N_\mathcal{X}$ is the number of chain iterations, $\mathcal{X}_t$ are the samples from sampling process and t is stage number. They start from an arbitrary state $\boldsymbol{x}_t^1 \in \mathcal{D}_{\boldsymbol{X}}$ that is updated sequentially based on a *transition kernel*. This kernel is to ensure that *Markov chains* fulfil *detailed balance* condition and are reversible, i.e., that the dynamics of the chains is not affected by the direction of iterative sampling. We omit an in-depth discussion of those properties and refer interested readers to exhaustive resources on the matter in Murphy (2012).

Nowadays, practitioners can choose from a vast of MCMC algorithms, ranging from the classical Metropolish-Hasting (MH), to advanced *Hamiltonian mechanics-inspired samplers*, *transitional MCMC algorithm* and *ensemble algorithms*. In the following, we will present two popular standard MCMC algorithms that are used in the context of Bayesian inference, but other alternative MCMC sampling methods can be also easily

implemented.

Metropolish-Hasting (MH)

The basic concept of MH is to choose a proposal distribution $q(\boldsymbol{x}_t^{s+1}|\boldsymbol{x}_t^s), s = 1, \cdots, N_{\mathcal{X}}$, between two adjacent samples from the current sample point $\boldsymbol{x}_t^s$ to the new point $\boldsymbol{x}_t^{s+1}$. If we have the first sample $\boldsymbol{x}_t^1$, we can generate the later samples for any number $N_{\mathcal{X}} > 1$. These samples can be summarized to describe the distribution of inversed parameters $\boldsymbol{x}_t$. For any given distribution $\pi(\boldsymbol{x}_t)$, we define the acceptance function, which is used to decide whether to accept this move, as follows:

$$f(\boldsymbol{x}_t^{s+1}|\boldsymbol{x}_t^s) = \min(1, \alpha) \tag{2.15}$$

$$\alpha = \min(1, \frac{q(\boldsymbol{x}_t^s|\boldsymbol{x}_t^{s+1})\pi(\boldsymbol{x}_t^{s+1}|\mathcal{Y}_t)}{q(\boldsymbol{x}_t^{s+1}|\boldsymbol{x}_t^s)\pi(\boldsymbol{x}_t^s|\mathcal{Y}_t)}) \tag{2.16}$$

In practice, in order to accept and reject the proposed candidates with probability in Equation (2.15) and Equation (2.16), a random variate u is sampled from a uniform distribution $U \sim \mathcal{U}(0, 1)$, and compares it to the probability α. If $\alpha > 1$, which implies $q(\boldsymbol{x}_t^s|\boldsymbol{x}_t^{s+1})\pi(\boldsymbol{x}_t^{s+1}|\mathcal{Y}_t) > q(\boldsymbol{x}_t^{s+1}|\boldsymbol{x}_t^s)\pi(\boldsymbol{x}_t^s|\mathcal{Y}_t)$, we accept the candidate sampled point $\boldsymbol{x}_t^{(\star)}$ from $q(\boldsymbol{x}_t^s|\boldsymbol{x}_t^{s+1})$. If $\alpha < 1$, we accept the candidate point $\boldsymbol{x}_t^{(\star)}$ with probability α. If the candidate is accepted, the new sample point $\boldsymbol{x}_t^{(\star)}$ is $\boldsymbol{x}_t^{s+1}$, otherwise stays the same as $\boldsymbol{x}_t^s$. A frequently employed proposal distribution is a Gaussian-type distribution centered at $\boldsymbol{x}_t^s$. The acceptance function can be modified as:

$$f(\boldsymbol{x}_t^{s+1}|\boldsymbol{x}_t^s) = \min(1, \alpha) = \min(1, \frac{q(\boldsymbol{x}_t^{s+1})}{q(\boldsymbol{x}_t^s)}) \tag{2.17}$$

The process of MH is shown in Algorithm 1. It is worth mentioning that although the proposal distribution $q(\boldsymbol{x}_t)$ can be any distribution, the closer it is to the actual target distribution $\pi(\boldsymbol{x}_t|\mathcal{Y})$, the more efficient the chain mixture would be. Also, we need to assign an initial position $\boldsymbol{x}_t^1 \in \mathcal{D}_{\boldsymbol{X}}$ that is not zero probability. Although the initial position does not affect the convergence of the sampling, a good initial guess would help to accelerate mixing for a Markov chain. With more paralleled N_{chain} chains and iterative

Algorithm 1: Metropolish-Hasting algorithm at t_{th} step

Data: $q(\boldsymbol{x}_t)$: Proposal distribution; $\pi(\boldsymbol{x}_t|\mathcal{Y})$: Target posterior.
Result: MCMC samples at t_{th} stage: $\mathcal{X}_t = \{\boldsymbol{x}_t^1, \cdots, \boldsymbol{x}_t^{N_\mathcal{X}}\}$

```
Initialization x_t^1 ∈ D_X;
for s ← 2 to N_X do
    Sample x_t^{s+1} ~ q(x_t^{s+1}|x_t^s);
    Compute acceptance probability α;
    Compute f(x_t^{s+1}|x_t^s) = min(1, α);
    Sample u ~ U(0, 1);
    Set candidate sample x_t^(⋆) to x_t^{s+1} with probability α;
end for
```

steps used ($\mathcal{X}_t = \{\mathcal{X}_{t\_1}, \cdots, \mathcal{X}_{t\_N_{chain}}\}$), where each chain $\mathcal{X}_{t\_i}$ contains $\{\boldsymbol{x}_{t\_i}^1, \cdots, \boldsymbol{x}_{t\_i}^{N_\mathcal{X}}\}$, the convergence will be improved accordingly. One schematic representation of the MH algorithm for sampling can be seen in Figure 2.6.

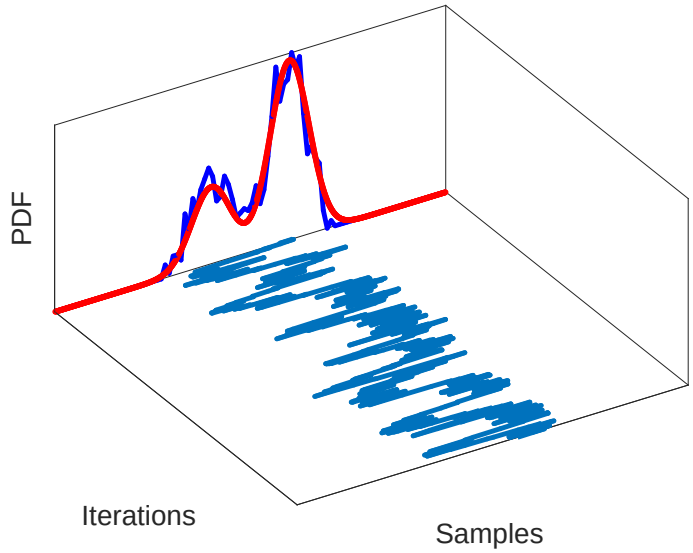


Figure 2.6: An schematic figure of the MH algorithm for sampling from a mixture 1D Gaussian at t_{th} stage

Affine invariant ensemble sampler (AIES)

Many MCMC algorithms perfom poorly when the target (i.e., posterior) distribution $\pi(\boldsymbol{x}_t|\mathcal{Y})$ exhibits strong, priorly unknown, inter-parameter correlation. For the MH, for example, it is challenging to construct a proposal distribution that propose states in the high-probability regions of the parameter space. This results in low acceptance rates which can only be improved by considerable tuning efforts. The AIES originally presented in Goodman & Weare (2010) alleviates this problem. Its invariance to affine

transformations means that if there exists an affine transformation of the *difficult-to-sample* (by standard MCMC methods) target distbution to an *easier-to-sample* target distribution, AIES samples both distributions equally well without explicitly requiring this affine transformation.

The AIES algorithm presented in Algorithm 2, simultaneously generates an *ensemble* of N_{chain} *Markov chains* at the t_{th} stage ($\mathcal{X}_t = \{\mathcal{X}_{t\_1}, \cdots, \mathcal{X}_{t\_N_{chain}}\}$). Each chain is called a *walker*. At the $\{s_{th}\}_{s\in\{1,\cdots,N_{\mathcal{X}}\}}$ iteration, all *Markov chain states* $\{\boldsymbol{x}^s_{t\_i}\}_{i\in\{1,\cdots,N_{chain}\}}$ are updated *walker by walker*. To update the i_{th} *walker*, the algorithm randomly picks a *conjugate walker* from $j \in \{1, \cdots, N_{chain}\}\backslash i$, i.e., excluding the current i_{th} walker. The *stretch move* is employed to generate proposals, thereby achieving the affine invariance property. This refers to proposing a new candidate along a straight line between the current walker and the *conjugate walker* with

$$\boldsymbol{x}^{(\star)}_t = \boldsymbol{x}^{(s)}_{t\_i} + z \cdot (\boldsymbol{x}^{(\tilde{s})}_{t\_j} - \boldsymbol{x}^{(s)}_{t\_i}) \tag{2.18}$$

where $\tilde{s} = s + 1$ if $j < i$ and $\tilde{t} = s$ otherwise, i.e., it denotes the latest available stage of the j_{th} *walker*. The stretch factor z is randomly drawn from the PDF

$$p(z|a) = \begin{cases} \frac{1}{\sqrt{z}(2\sqrt{a} - \frac{2}{\sqrt{a}})} & \text{if } z \in [1/a, a] \\ 0 & \text{otherwise} \end{cases} \tag{2.19}$$

which relies on the tuning parameter $a > 1$. The candidate sample $x^{(\star)}_t \in \mathcal{D}_{\boldsymbol{X}} \subseteq \mathbb{R}^M$ is then accepted as the new location of the i_{th} walker with probability:

$$\alpha = \min(1, z^{M-1} \frac{\pi(x^{(\star)}_t|\mathcal{Y})}{\pi(x^{(s)}_{t\_i}|\mathcal{Y})}) \tag{2.20}$$

This is repeated for all N_{chain} walkers in the ensemble. The resulting chains fulfil the *detailed balanced condition*. A practical advantage of the AIES algorithm is that it only has a single scalar tuning parameter a, which is often set to $a = 2$ (Wagner et al., 2022).

Algorithm 2: Affine invariant ensemble sampler algorithm at t_{th} step

Data: $\pi(\boldsymbol{x}_t|\mathcal{Y})$: Target posterior; tuning parameter a

Result: MCMC samples at t_{th} stage: $\mathcal{X}_t = \{\mathcal{X}_{t\_1}, \cdots, \mathcal{X}_{t\_N_{chain}}\})$, with $\mathcal{X}_{t\_i} = \{\boldsymbol{x}_{t\_i}^1, \cdots, \boldsymbol{x}_{t\_i}^{N_\mathcal{X}}\}$

Initialization N_{chain} samples $\{\boldsymbol{x}_{t\_1}^1, \cdots, \boldsymbol{x}_{t\_N_{chain}}^1\}$, with $\boldsymbol{x}_{t\_i}^1 \in \mathcal{D}_{\boldsymbol{X}}$

for $s \leftarrow 2$ *to* $N_\mathcal{X}$ **do**

  for $i \in \{1, \cdots, N_{chain}\}$ **do**

    Pick random j from $\{1, \cdots, N_{chain}\} \backslash i$;

    Propose $x_t^{(\star)}$ with Equation (2.18);

    Set $\boldsymbol{x}_{t\_i}^s = x_t^{(\star)}$ with probability α (see Equation (2.20));

  end for

end for

2.4.5 Sequential Monte Carlo

Like MCMC, Sequential Monte Carlo (SMC), is another popular sampling method for a dynamic model. The parameters $\boldsymbol{x_t}$ which we are interested can be linked with the observations $\mathcal{Y}_t$ in a time series way:

$$\begin{aligned} \boldsymbol{x_t} &= g(\boldsymbol{x_{t-1}}) + \boldsymbol{v} \qquad \text{(state equation)} \\ \mathcal{Y}_t &= m(\boldsymbol{x_t}) + \boldsymbol{w} \qquad \text{(observation equation)} \end{aligned} \tag{2.21}$$

which stands for the prediction step and correction step, respectively. $g(\cdot)$ and $m(\cdot)$ denote *transition equation* and *emission equation*, respectively. These functions can be either linear or nonlinear. $\boldsymbol{v}$ and $\boldsymbol{w}$ are independent random variables (or vectors) that denote process noise and observation noise, respectively.

In this setting, we only focus on nonlinear non-Gaussian dynamic model, often referred to as a *particle filter*. The feature of the non-linearity is used to relax the constraints at some simplified models (i.e., Kalman filter (KF) or Unscented Kalman filter (UKF), etc.). SMC, a sequential Bayesian technique, involves getting a population of non-correlated particles where each sample denotes to a potential parameter set. Subsequently, these particles undergo filtering to approximate the final distribution in a manner that aligns with the posterior distribution $\pi(\boldsymbol{x}_t|\mathcal{Y})$. The filtering process usually encompasses several steps: initial samples, reweighting and resampling, which are sequentially applied while new observations $\mathcal{Y}_t$ are added. There are different flavours of *particle fil-*

ter. For book-length treatment, see Murphy (2012). Due to space limitation, we will only present two standard SMC algorithms that are used in the context of Bayesian inference, but other effective *particle filter* methods can be also easily implemented.

Sequential importance sampling

The concept is to estimate the belief state of the trajectory stage by utilizing a weighted set of particles:

$$\pi(\boldsymbol{x}_{1:t}|\mathcal{Y}_{1:t}) \approx \sum_{s=1}^{S} \tilde{w}_t^s \delta_{\boldsymbol{x}_{1:t}^s}(\boldsymbol{x}_{1:t}) \tag{2.22}$$

$$\tilde{w}_t^s = \frac{w_t^s}{\sum_{s=1}^{S}(w_t^s)} \tag{2.23}$$

where $\tilde{w}_t^s$ represents the normalised weight of sample $s = \{1, \cdots, N\}$ at time t and δ denotes the *Dirac delta function*. Utilizing this representation, we can straightforwardly calculate the marginal distribution over the most recent state, $\pi(\boldsymbol{x}_t|\mathcal{Y}_{1:t})$, by simply disregarding the preceding segments of the trajectory $\boldsymbol{x}_{1:t}$.

We update this belief state through importance sampling. If the proposal takes the form $q(\boldsymbol{x}_{1:t}|\mathcal{Y}_{1:t})$, we can recursively express the numerator as follows:

$$\pi(\boldsymbol{x}_{1:t}|\mathcal{Y}_{1:t}) \propto \pi(\mathcal{Y}_t|\boldsymbol{x}_t)\pi(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})\pi(\boldsymbol{x}_{t-1}|\mathcal{Y}_{t-1}) \tag{2.24}$$

where we have assumed the customary Markov assumptions. Our focus will be on proposal densities in the following form:

$$\pi(\boldsymbol{x}_{1:t}|\mathcal{Y}_{1:t}) = \pi(\boldsymbol{x}_t|\boldsymbol{x}_{1:t-1}, \mathcal{Y}_{1:t})\pi(\boldsymbol{x}_{1:t-1}|\mathcal{Y}_{1:t-1}) \tag{2.25}$$

Next, incorporating the new state $\boldsymbol{x}_t$ at the end, we assume the necessity to retain only the most recent portion of the trajectory and observation sequence, rather than the entire history $1:t$ for simplicity. In such a scenario, the weight takes the following form:

$$w_t^s \propto w_{t-1}^s \frac{\pi(\mathcal{Y}_t|\boldsymbol{x}_t^s)\pi(\boldsymbol{x}_t^s|\boldsymbol{x}_{t-1}^s)}{q(\boldsymbol{x}_t^s|\boldsymbol{x}_{t-1}^s, \mathcal{Y}_t)} \tag{2.26}$$

Hence we can approximate the posterior $\pi(\boldsymbol{x}_t|\mathcal{Y}_{1:t})$ filtered density using Equation (2.27)

$$\pi(\boldsymbol{x}_t|\mathcal{Y}_{1:t}) \approx \sum_{s=1}^{S} \tilde{w}_t^s \delta_{\boldsymbol{x}_t^s}(\boldsymbol{x}_t) \tag{2.27}$$

As $S \to \infty$, it can be demonstrated that this converges to the true posterior. All initial particles share the same weight values and can be uniformly sampled using *Latin hypercube sampling* within feasible ranges of parameter values. The whole SIS without resampling process is shown in Figure 2.7 and SIS algorithm is presented in Algorithm 3.

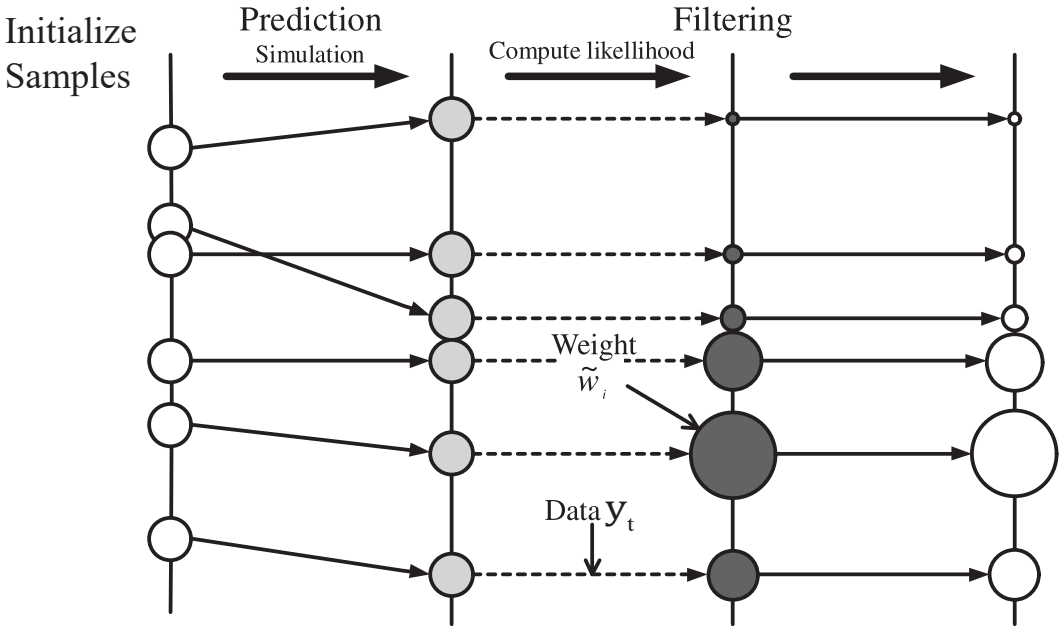


Figure 2.7: Particle filtering using SIS adapted from Nguyen & Nestorović (2016)

Algorithm 3: Sequential importance sampling algorithm at t_{th} step

Data: Samples $\boldsymbol{x}_{t-1}^s$ with weights w_{t-1}^s, $s = \{1, \cdots, N\}$; observation $\mathcal{Y}_t$ at t_{th} stage
Result: SMC samples with normalized weights $\tilde{w}_t^s$ at t_{th} stage:
$\boldsymbol{x}_t^{(\star)} = \mathcal{X}_t = \{\boldsymbol{x}_t^1, \cdots, \boldsymbol{x}_t^N\}$)

```
for s ← 1 to N do
    Sample from proposal distribution x_t^s ~ q(x_t^s | x_{t-1}^s, Y_t);
    Compute weight using Equation (2.26);
end for
Normalized weights;
```

Sequential importance sampling and resampling

The fundamental SIS algorithm encounters challenges after a few steps due to the fact that the majority of the particles will possess negligible weight. This issue is referred to as the *degeneracy problem*, particularly prevalent when sampling in a high-dimensional space. We can measure the degree of degeneracy using the *effective sample size*, defined as:

$$\hat{S}_{eff} = \frac{1}{\sum_{s=1}^{S}(w_t^s)^2} \tag{2.28}$$

If the variance of the weights is substantial, we are essentially allocating resources to update particles with low weight, which contribute minimally to our posterior estimate. An effective approach to address the degeneracy problem involves introducing a resampling step. This enhancement to the basic SIS algorithm monitors the *effective sample size*. Whenever it falls below a threshold S_{min}, particles with low weight are eliminated, and replicates of the surviving particles are generated. This procedure is illustrated in Figure 2.8 and the comprehensive SISR algorithm is outlined in Algorithm 4. In this case, $\boldsymbol{x}_t^{(\star)} \neq \mathcal{X}_t$ comes from fact that the resampling process discards low weighted particles.

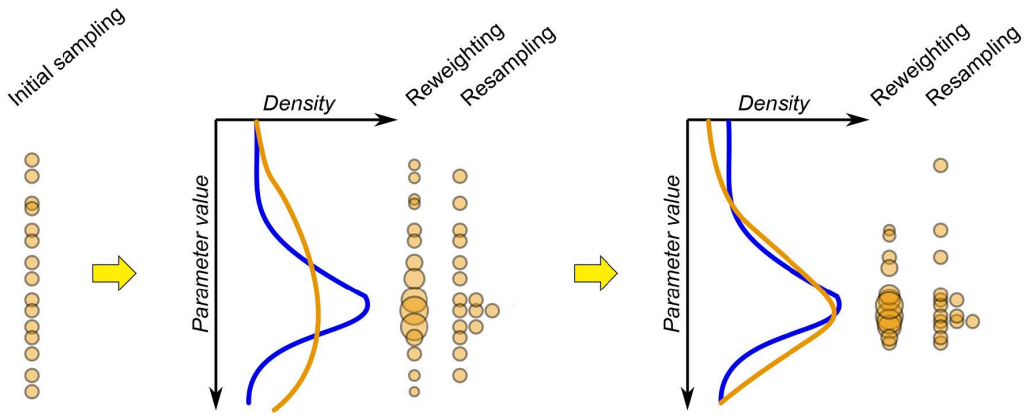


Figure 2.8: Principle of SISR algorithm adapted from Speich et al. (2021)

2.5 Choices for sampling methods

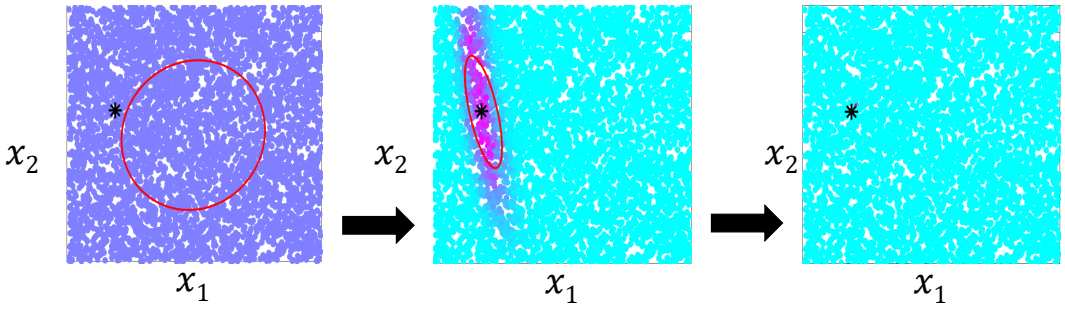
In a continuous Bayesian calibration process, MCMC and SMC are two prevalent sampling methods. They are differed based on different sampling principles. MCMC-derived uncertain variables are iteratively constructed based on *random walk* and out-

Algorithm 4: Sequential importance sampling and resampling algorithm at t_{th} step

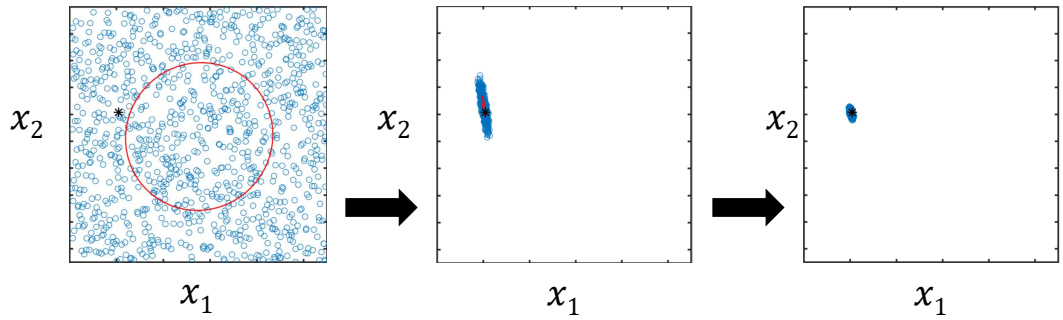
Data: Samples $\boldsymbol{x}_{t-1}^s$ with weights w_{t-1}^s, $s = \{1, \cdots, N\}$; observation $\mathcal{Y}_t$ at t_{th} stage
Result: SMC samples with normalized weights $\tilde{w}_t^s$ at t_{th} stage:
$\boldsymbol{x}_t^{(\star)} = \{\boldsymbol{x}_t^1, \cdots, \boldsymbol{x}_t^N\}$)

```
for s ← 1 to N do
    Sample from proposal distribution x_t^s ~ q(x_t^s | x_{t-1}^s, Y_t);
    Compute weight using Equation (2.26);
end for
Normalized weights;
Calculate degeneracy measure using Equation (2.28);
if Ŝ_eff < S then
    Resample;
end if
```

performs in high dimensions, while SMC-derived uncertain variables are represented by updating weights on the particles and yields faster inference results. As shown in Figures 2.9a and 2.9b, both methods demonstrates their effectiveness in updating parameters in a two dimensional space example.



(a) Particle filter with Sequential importance sampling



(b) Markov Chain Monte Carlo using Affine invariant ensemble sampler

Figure 2.9: SMC and MCMC in a two-dimensional example

SMC, with its non-iterative nature, offered the advantage of parallel processing, significantly enhancing computational efficiency. This attribute makes SMC a favorable option when quick results are imperative. However, *particle filter* doesnot work well in

high dimensional spaces (Murphy, 2012). Instead, MCMC, with its random walk approach, showcased its proficiency in exploring high-dimensional parameter spaces, making it a suitable choice for complex and multifaceted problems. Therefore, the choice between these methods should be guided by the specific needs and priorities of the analysis, whether precision or computational efficiency takes precedence. Since geotechnical engineering are naturally in high dimensional space, this thesis adopts MCMC during the sequential Bayesian calibration for the next sampling analysis.

Chapter 3

Uncertainty quantification in high dimensions

Uncertainty quantification (UQ) endeavors to consider the uncertainties associated with the parameters in the model of a physical system and examine their influence on the system response. A well-established representation of an uncertainty quantification problem is outlined next. The premise is that any such problem can be depicted as a combination of these ingredients, as illustrated in Figure 3.1:

- A computational model $\mathcal{M}$. This encompasses a broad spectrum, ranging from an analytical function in its simplest form to a black box housing various levels of partial differential equations (e.g., finite element packages or finite difference packages). Generally, the computational model $\mathcal{M}$ establishes a mapping from a set of input parameters $\boldsymbol{x}$) to one or more Quantities of interest (QoI), often denoted as *model responses*.

- The sources of uncertainties in input space. This step involves identifying the input parameters that are uncertain and describing them within a probabilistic context.

- Uncertainty propagation from input parameters $\boldsymbol{x}$ to $\boldsymbol{y}$. This step pertains to quantifying the QoI by propagating the uncertainty of the input space through the computational model $\mathcal{M}$.

- Iterative updating of the source of uncertainty. This step encompasses various techniques employed to refine the information related to the identified sources of uncertainty above. Examples includes *sensitivity analysis* or *Bayesian inference*. If we are using *sensitivity analysis* to update the uncertainties, it can be explicitly called as *forward problems*. If we are using *Bayesian inference* to update the uncertainties, it can be called as *inverse problems*.

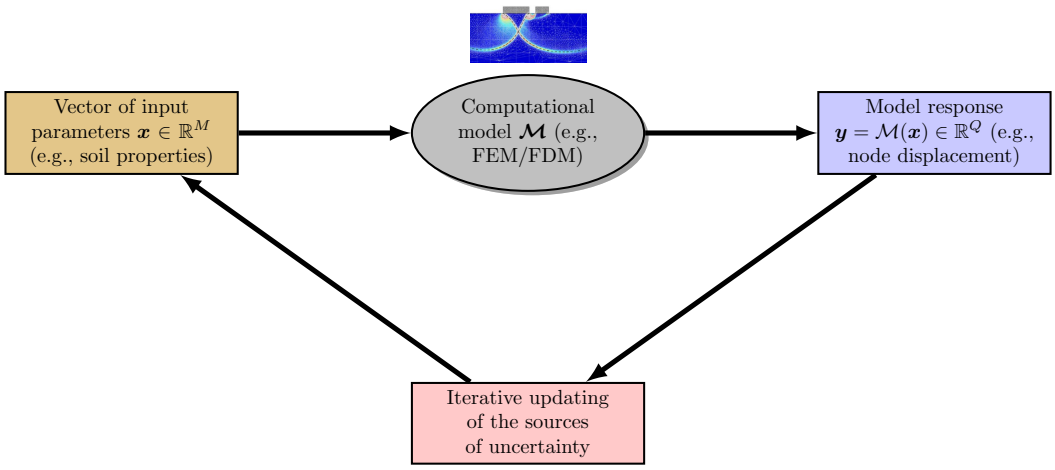


Figure 3.1: Global framework for uncertainty quantification

3.1 Problem statement

In modern engineering applications, uncertainty quantification (UQ) often involves simulations with a large number of input parameters. The computational model is frequently treated as a black box, where only the input parameters $\boldsymbol{x}$ and the corresponding model response $\boldsymbol{y}$ are available. Addressing uncertainties through traditional *Monte Carlo* methods in such systems can be computationally intensive, posing a significant challenge. To alleviate this computational burden, surrogate models $\tilde{\mathcal{M}}$, illustrated in Figure 3.2, can be employed. The solid line denotes the current working flow. These surrogate models approximate the original model with a cost-effective replacement model, allowing for efficient exploration of the input parameter space. A surrogate model $\tilde{\mathcal{M}}$ can be then expressed as:

$$\tilde{\mathcal{M}}(\boldsymbol{X}) \overset{\text{def}}{=} \mathcal{M}(\boldsymbol{X}) - \mathcal{R}(\boldsymbol{X}) \tag{3.1}$$

where $\mathcal{R}$ is the residual between the original model and the surrogate. The overarching concept of Equation (3.1) is straightforward: working with a finite set of input parameters and their corresponding model outputs, commonly referred to as an *experiment of design*.

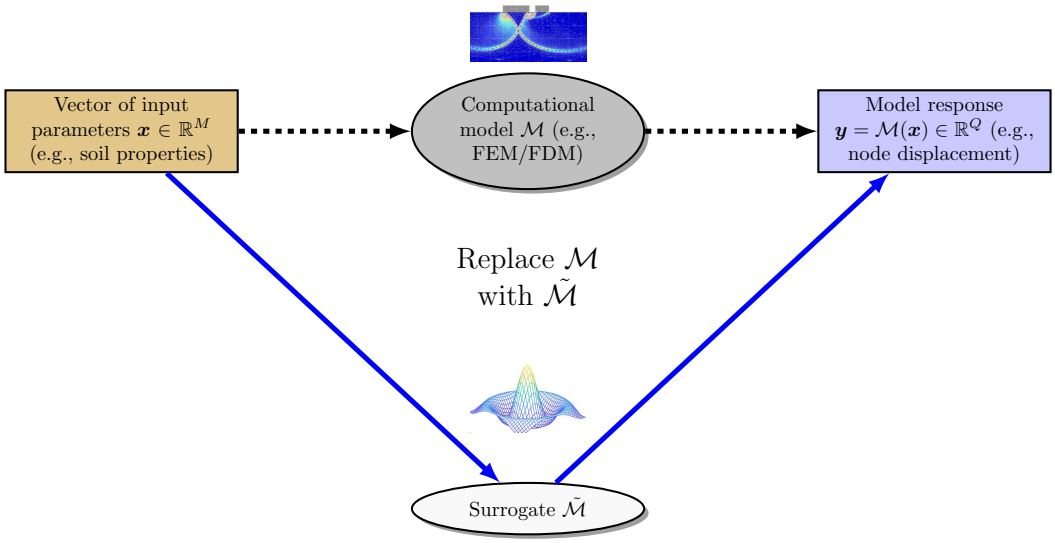


Figure 3.2: Using a surrogate to obtain the model response

In high dimensions scenarios, surrogate models tend to exhibit diminished performance, coupled with escalating computational and storage costs-a challenge commonly acknowledged as the *curse of dimensionality*. Consequently, challenges may arise from:

(1) in large-scale scenarios, the computational intractability of conducting thousands of forward simulations is a common challenge; (2) the complexity of the input/output space dimensions adds another layer of difficulty;(3) the large dimensionality of the input/output space complicates the sampling process; Thus, to alleviate this in a statistical inverse problems, methods can be broadly categorized in three groups: (1) choose an appropriate type of surrogates to accelerate a forward simulation; (2) reduce the size of the input/output space, i.e., sensitive analysis or Dimensionality reduction techniques; (3) sequential efficient sampling method in *experiment of design* (*active learning*) or in a posterior (e.g., MCMC).

Table 3.1: Surrogate model choices

Name	Shape	Parameters
Polynomial chaos expansions	$\tilde{\mathcal{M}}(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}\in\mathcal{A}} \boldsymbol{y}_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$	$\boldsymbol{y}_{\boldsymbol{\alpha}}$
Low-rank tensor approximations	$\tilde{\mathcal{M}}(\boldsymbol{x}) = \sum_{l=1}^{R} b_l \left(\prod_{i=1}^{M} v_l^i x_i\right)$	$b_l,\ z_{k,l}^i$
Kriging (a.k.a Gaussian processs)	$\tilde{\mathcal{M}}(\boldsymbol{x}) = \boldsymbol{\beta}^T \cdot \boldsymbol{f}(\boldsymbol{x}) + Z(\boldsymbol{x}, \omega)$	$\boldsymbol{\beta},\ \sigma_Z^2,\ \boldsymbol{\theta}$
Support vector machines	$\tilde{\mathcal{M}}(\boldsymbol{x}) = \sum_{i=1}^{m} a_i K(\boldsymbol{x}_i, \boldsymbol{x}) + b$	$\boldsymbol{a},\ b$
Neural networks	$\tilde{\mathcal{M}}(\boldsymbol{x}) = f_n\left(\cdots f_2(b_2 + f_1(b_1 + \boldsymbol{w}_1 \cdot \boldsymbol{x}) \cdot \boldsymbol{w}_2)\right)$	$\boldsymbol{w},\ \boldsymbol{b}$

3.2 Surrogate model choices

Ongoing research on surrogate modelling focuses on various problems, like hyperparametes tuning, mathematical explanation or the accuracy based on few observations. A range of techniques exist in surrogate modelling, in which some of them are listed in the table 3.1. Among these popular surrogate models above, it is crucial to choose the most appropriate model for our own geotechnical problem. Inspired by Torre et al. (2019), we only focus on Polynomial chaos expansion (PCE) which is presented next. Compared with other surrogate models based on benchmark data sets (e.g., *Ishigami function*, *23-bar horizontal truss*), PCE exhibits additional advantageous properties, such as:

- PCE excels in various tasks, necessitating only minimal parameters tuning for adaptation to specified data considerations.
- PCE not only provides precise point-wise predictions for the output at given inputs, but also furnishes relevant statistics in the presence of input uncertainties. This benefits in UQ process as it results from the integration of the PCE model with a suitable probabilistic characterisation of the input model.
- The coefficients of PCE makes it free to get sensitivity analysis results.
- The output produced by PCE is analytically expressed as a simple polynomial of the input. This characteristic makes the model easy to interpret from a mathematical standpoint. This is in contrast with black-box models (e.g., neural networks) or classification-based models (e.g., tree models), where our reliance is placed on the assumption that training data is sufficiently big to yield an accurate surrogate.

- PCE achieves acceptable performance levels with a relatively modest amount of data in the *experiment of design*.

Note: The choice for PCE in our thesis does not guarantee the conclusion that PCE outperforms than other surrogate models. The properties obtained above are only based on some simple toy problems. Therefore, a choice for surrogate models is problem-specified and it is not always the same case when it encounters into different geotechnical problems. However, the PCE advantages above (Torre et al., 2019) indeed give us confidence to get started with PCE first. But other advanced surrogate models can be also easily employed.

3.3 Polynomial chaos expansion

Polynomial chaos expansions (PCE) represent a potent metamodelling technique designed to provide a functional approximation of a computational model. This approximation is achieved through its spectral representation of the model on a carefully built basis of polynomial functions. Consider a random vector with independent components $\boldsymbol{X} \in \mathbb{R}^M$ characterized by the joint PDF $f_{\boldsymbol{X}}$. Additionally, suppose there exists a computation model with finite variance, defined as a mapping $Y = \mathcal{M}(\boldsymbol{X})$, with $Y \in \mathbb{R}$. This can be expressed as follows:

$$\mathbb{E}[Y^2] = \int_{\mathcal{D}_{\boldsymbol{X}}} \mathcal{M}^2(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} < \infty \tag{3.2}$$

Then, a PCE $\mathcal{M}(\boldsymbol{X})$ can be represented as:

$$Y = \mathcal{M}(\boldsymbol{X}) \approx \sum_{\alpha \in \mathcal{A}} \boldsymbol{y}_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{X}) \tag{3.3}$$

where $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{X})$ denotes multivariate polynomials orthonormal relative to $f_{\boldsymbol{X}}$, The multi-index $\boldsymbol{\alpha} \in \mathcal{A}$ identifies the truncated multivariate polynomials $\Psi_{\boldsymbol{\alpha}}$. $\boldsymbol{y}_{\boldsymbol{\alpha}} \in \mathbb{R}$ corresponds to the corresponding coefficients. There are four main ingredients for a PCE: (1) Construct the basis functions $\Psi_{\boldsymbol{\alpha}}$; (2) Compute the coefficients $\boldsymbol{y}_{\boldsymbol{\alpha}}$; (3) Evaluate the precision of the PCE; (4) Post process the PCE. A visualized PCE example with a two-dimensional

input can be seen in Figure 3.3.

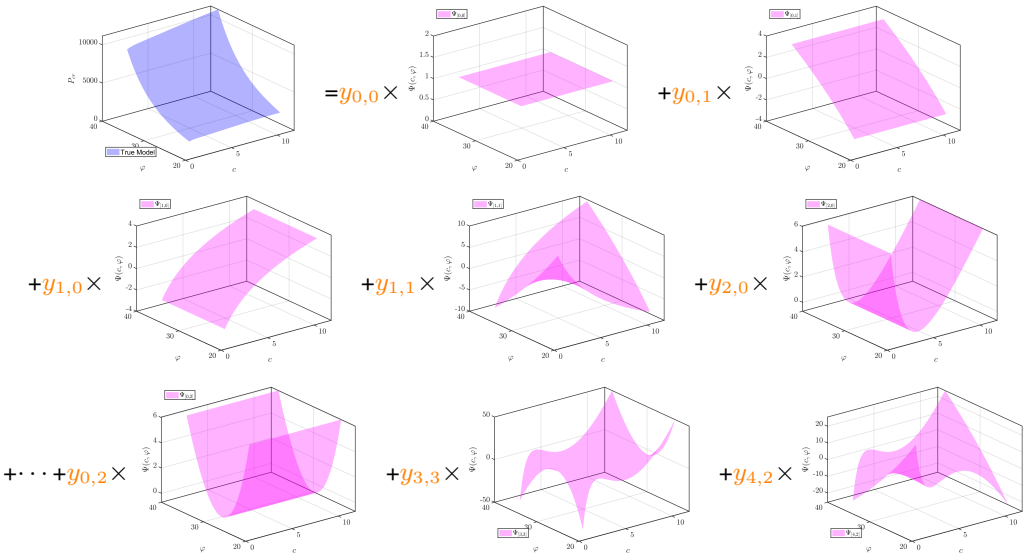


Figure 3.3: PCE visualization in a two-dimensional example from Sudret (2019)

3.3.1 Construct the basis functions Ψ_α

The polynomial basis $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$ in Equation (3.3) is conventionally constructed using a set of *univariate orthonormal polynomials* $\phi_k^i(x_i)$, where these polynomials satisfy:

$$\left\langle \phi_j^i, \phi_k^i \right\rangle = \delta_{jk} \tag{3.4}$$

δ_{jk} is the *Kronecker symbol*. The multivariate polynomials Ψ_α are then constructed by taking the tensor product of their univariate counterparts:

$$\Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \overset{\text{def}}{=} \prod_{i=1}^{M} \phi_{\alpha_i}^i(x_i) \tag{3.5}$$

in which, $\boldsymbol{\alpha} = \{\alpha_1, \cdots, \alpha_M\}$ $\alpha_i \in \mathbb{N}^M$ of degree $|\boldsymbol{\alpha}| = \sum_{i=1}^M \alpha_i$, and $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$ satisfies orthonormal properties as: $\langle \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}), \Psi_{\boldsymbol{\beta}}(\boldsymbol{x}) \rangle = \delta_{\boldsymbol{\alpha\beta}}$, where the symbol $\delta_{\boldsymbol{\alpha\beta}}$ is an extension *Kronecker symbol* to the multi-dimensional case. Classical families of orthogonal polynomials have been discovered historically in table 3.2.

$$A^{M,p,q} = \{\alpha \in A^{M,p} : ||\alpha||_q \leq p\}; \text{ where } ||\alpha||_q = \left(\sum_{i=1}^{M} \alpha_i^q \right)^{1/q} \tag{3.6}$$

Table 3.2: Univariate orthogonal polynomials

Name	**Support $\mathcal{D}_X$**	**Weight function f_X**
Legendre	$[a; b]$	uniform
Jacobi	$[a; b]$	Beta
Hermite	$(-\infty; \infty)$	Gaussian
Laguerre	$[; \infty)$	Gamma

The infinite series expansion cannot be handled in practical computations. Two truncation schemes can be applied: (1) restriction of maximum interaction p; (2) reduce hyperbolic truncation q. An example of the truncation in two dimensions for various values of p and q is depicted in Figure 3.4.

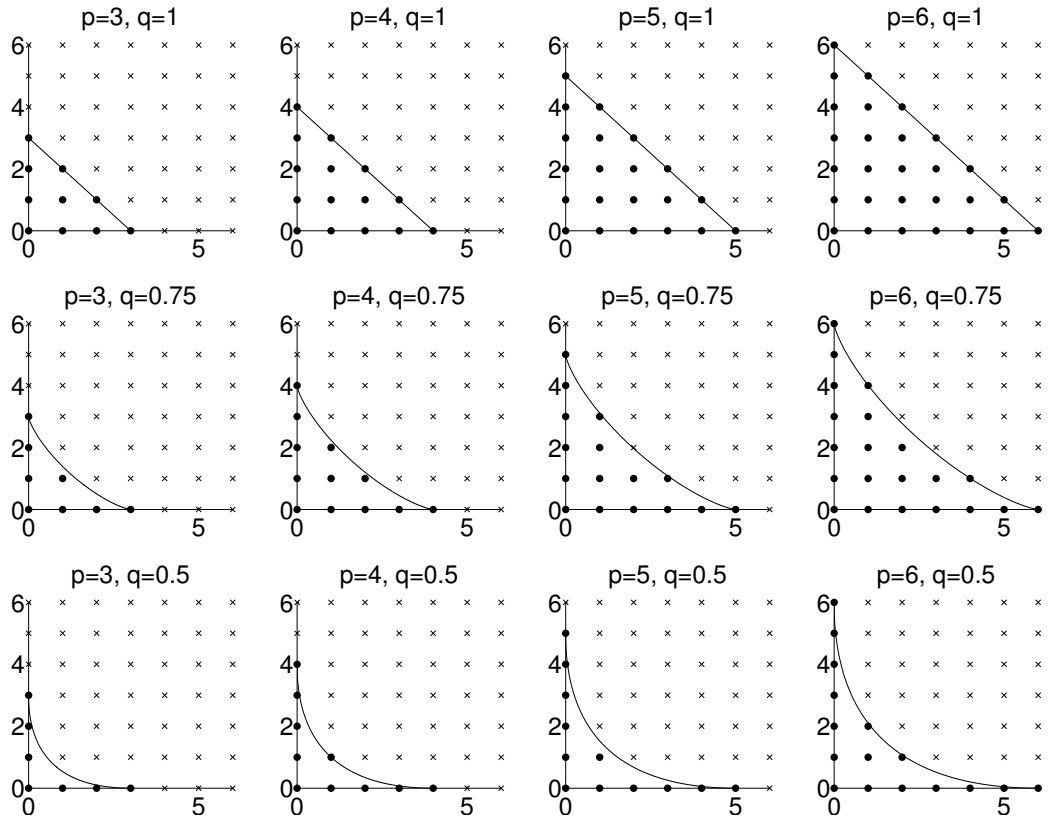


Figure 3.4: Truncation set for varying values of p and q from Wagner et al. (2022)

3.3.2 Compute the coefficients y_α

Various methods are available for computing the coefficients $\boldsymbol{y_\alpha}$ of the PCE with a known basis. Two primary strategies commonly employed are *projection* and *regression*. *Projection* methods leverage the orthogonality of the basis to determine the coefficients. On the other hand, *regression* methods utilize standard linear regression approaches to

solve the system. Below are some of the methods:

Projection method

The evaluation of the polynomial coefficients $\boldsymbol{y}_{\boldsymbol{\alpha}}$ can be reformulated as the calculation of the expectation value in

$$\boldsymbol{y}_{\boldsymbol{\alpha}} = E[\Psi_{\boldsymbol{\alpha}}(\boldsymbol{X}) \cdot \mathcal{M}(\boldsymbol{X})] \tag{3.7}$$

Least-squares minimization

This approach casts Equation (3.7) as a least-squares minimisation problem as shown:

$$\hat{\boldsymbol{y}}_{\boldsymbol{\alpha}} = \text{argmin}\mathbb{E}[\varepsilon_P^2(\boldsymbol{X})] = \text{argmin}\mathbb{E}[(\tilde{\mathcal{M}}(\boldsymbol{X}) - \mathcal{M}(\boldsymbol{X}))^2] \tag{3.8}$$

Some other regression methods can be also easily implemented, i.e., *ordinary least square*, *compressive sensing* or *stochastic collocation*.

3.3.3 Check the accuracy of the PCE

Error estimators are essential for quantifying the fidelity of a surrogate model in approximating the original model. To determine the precision of a surrogate, *leve-one-out* (LOO) is a good way to assess the cross validation error. For a PCE consisting N metamodels, error ε_{LOO} can be shown as:

$$\varepsilon_{\text{LOO}} = \frac{1}{N} \sum_{i=1}^{N} \left(\mathcal{M}(\boldsymbol{x}^{(i)}) - \tilde{\mathcal{M}}(\boldsymbol{x}^{(i)}) \right)^2 \tag{3.9}$$

3.3.4 Post process the PCE

Due to the inherent orthogonality with the polynomial basis functions, we can derive explicit expressions for moments characterizing the model output. The first pair moments of a PCE find their manifestation within is coefficients. Specifically, the mean

and variance of a PCE manifest as:

$$E[Y] = \mathbb{E}[\tilde{\mathcal{M}}(\boldsymbol{X})] = y_{\mathbf{0}} \tag{3.10}$$

$$\mathrm{Var}[Y] = E[Y^2] - E[Y]^2 \approx \sum_{\alpha \in A \setminus \{0\}} y_\alpha^2 \tag{3.11}$$

Another crucial aspect of PCE post-processing lies in the fact the coefficients encapsulate significant details regarding the *ANOVA* decomposition of a surrogate. This information can be harnessed to efficiently compute global sensitivity at a minimal computational expense. Because it is popular to make choices on the most important variables and update the uncertainties based on the sensitivity results. Thus, strictly speaking, *sensitivity analysis* also belongs to the category of DR.

3.4 Dimensionality reduction

Geotechnical problems inherently involve high dimensionality, posing challenges for learning methods like surrogate modeling. Technical constraints impact the storage and processing of such huge amount of data. Furthermore, as input and output data expand, independent scalar surrogate models show inadequate in accurately capturing the covariance matrix of the original data, leading to less reliable predictions. Consequently, in high dimensional space, the need for Dimensionality reduction (DR) becomes more critical.

The transformation from the *original space* $\mathcal{D}_{\boldsymbol{Y}} \subseteq \mathbb{R}^N$ to a *reduced space* $\mathcal{D}_{\boldsymbol{Z}} \subseteq \mathbb{R}^n$ with $n \ll N$ is the general form of a DR mapping:

$$\mathcal{T}_{DR} : \mathcal{D}_{\boldsymbol{Y}} \rightarrow \mathcal{D}_{\boldsymbol{Z}} \tag{3.12}$$

where the underlying assumption is that $\mathcal{D}_{\boldsymbol{Z}}$ is embedded inside $\mathcal{D}_{\boldsymbol{Y}}$. There exists a large set of DR techniques ranging from linear to nonlinear approaches. A simple but effective linear technique, that is widely used today, is the *principal component analysis* (PCA). It is popular across a wide range of disciplines and is closely related to the *Karhunen-Loève expansion* and *proper orthogonal decomposition*. In practical applications, the PCA process is involves the estimation of the expectation $\boldsymbol{\mu}_{\boldsymbol{Y}} \approx \mathbb{E}[\boldsymbol{Y}]$ and the covariance matrix

$\Sigma_{\boldsymbol{Y}} \approx Cov[\boldsymbol{Y}]$. The N eigenvectors of this covariance matrix are represented by ϕ_p for $p = 1, \cdots, N$. The corresponding eigenvalue λ_p signifies the variance of $\boldsymbol{Y}$ in direction of the $p-th$ principle component. Consequently, the random vector $\boldsymbol{Y}$ can be expressed through its n principal components with the highest variance:

$$\boldsymbol{Y} \approx \boldsymbol{Y}^{PCA} = \boldsymbol{\mu_Y} + \sum_{p=1}^{n} z_p \phi_p \tag{3.13}$$

The selection of the number n is determined such that $\sum_{p=1}^{n} \lambda_p = (1-\varepsilon_0) \sum_{p=1}^{N} \lambda_p$, where ε_0 typically set to 0.01. Consequently, the model output $\boldsymbol{Y} \in \mathbb{R}^N$ can be processed by a linear transformation of the principle component vector $\boldsymbol{Z} = (z_1, \cdots, z_n)$. This reduction in the dimensionality, from N to n $n \ll N$, is then achieved. When data sets become more complex, more advanced DR techniques can be used, such as *multi-dimensional scaling*, *kernel principal component analysis* and *auto-encoder*.

3.5 DR-based surrogate in Bayesian inference

In high dimensions, a popular two-step approach is frequently adopted to address such challenges: first, the input/output dimensions are reduced; subsequently, the surrogate model is directly constructed within the *reduced space*. Performance of DR-based surrogate has showed superior performance in some benchmarks (Lataniotis, 2019). Take PCA and PCE as a DR example in Figure 3.5, the amalgamation of PCA-PCE constitutes an efficient surrogate modelling technique. The solid line denotes the current working flow.

When it comes to Bayesian inference process accelerated by DR-based surrogate, working principles can be shown in Figure 3.6 as below:

- Step one: PCA technique reduces the size of input/output into principal components.
- Step two: construct PCE based on the principal components from step one.
- Step three: Pass obtained PCE into Bayesian inference process and evaluate the likelihood.

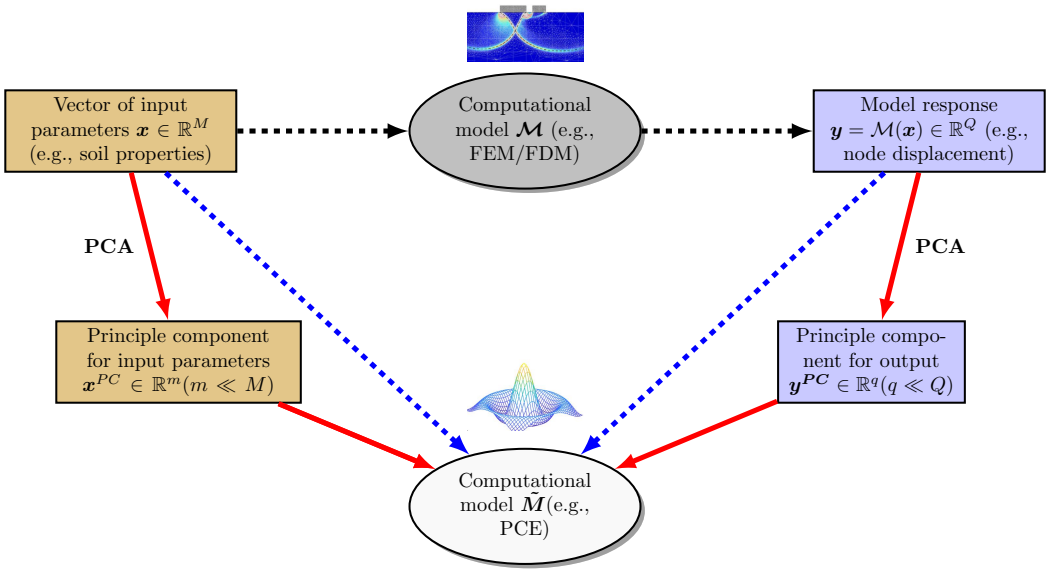


Figure 3.5: Dimensionality reduction (DR) schematic example

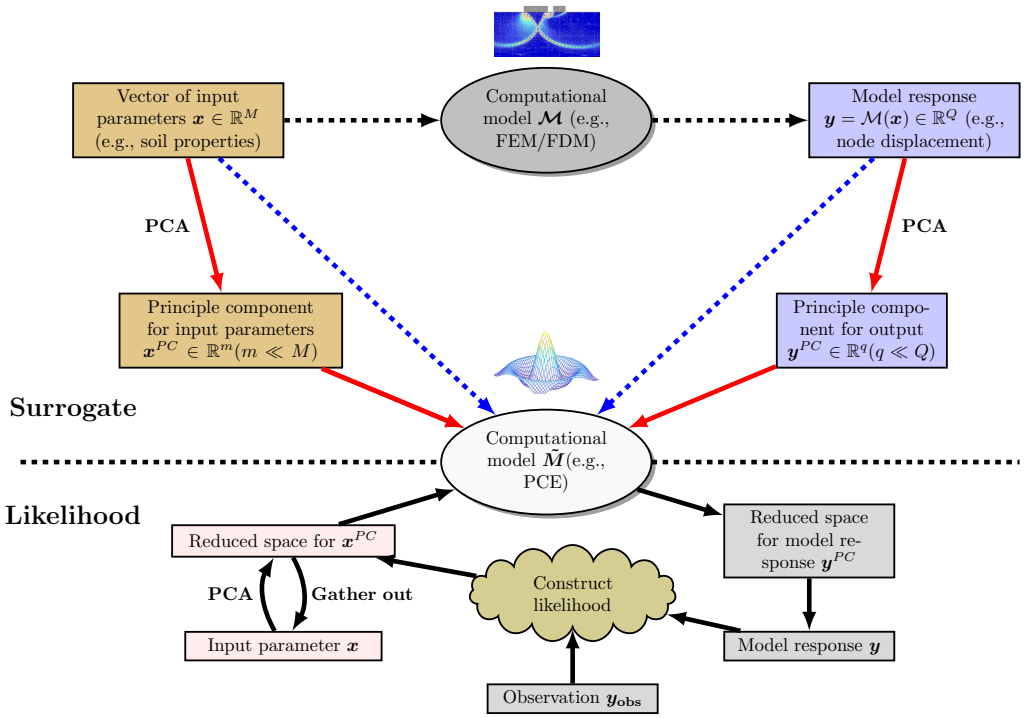


Figure 3.6: Bayesian inference accelerated by a PCA-PCE surrogate

Note: DR-based surrogates are not only limited to Bayesian inference in high dimensions. Even in low dimension's inference, DR-based surrogates still outperform than the case with surrogate models without any DR. Because DR is more than a data compressing technique in Bayesian inference, it can capture the covariance matrix of the original data to make more accurate predictions. Due to this feature, compared to the traditional scalar models (e.g., PCE, kriging), DR-based surrogates make it possible to consider multiple

outputs predictions.

Chapter 4

Next plans

4.1 Active learning for a surrogate

In high dimensions, *experiment of design* based on *Monte Carlo* for a surrogate can be huge and the computational cost is too expensive. Active learning provides as a tool to explore the random space in a sequential and much affordable way. They are built adaptively in an attempt to finely approximate the limit-state surface in areas of high probability mass. Next, we hope to identify new sampling locations by minimizing an acquisition function based on a PCE model.

4.2 A unified and scalable digital twin for piles

Urgent need for a unified and scalable digital twin

To address these challenges in real time, digital twin (DT) has gained popularity in handling abundant data and predict the pile's response in a more organized and accurate manner (Wang et al., 2021). DT can effectively leverage various data sources, including physical models, sensor updates and operating history. They seamlessly integrate simulation processes to dynamically replicate the behavior of a physical system in a virtual space in real-time.

However, in a context of geotechnical engineering problem, while the value propo-

sition of digital twin has become widely appreciated, the pile design process remains in a custom production phase. Current digital twin for offshore piles are still bespoken, relying on highly specialized implementations and thus requiring considerable resources and expertise to deploy and maintain. Therefore, it is necessary to move toward digital twins at scale by developing a rigorous and unified mathematical foundation.

Benefits of Partially observed probabilistic graphical model

Introducing probabilistic graphical model (PGM) involving in Bayesian inverse analysis can visualize the calculation and combine decision theory naturally to make prompt predictions. Besides, as a mathematical and rigorous foundation, PGM is proposed to support the transition from custom defined model towards accessible digital twins at scale. Based on such flexible asset-specific models, the entire loading life-cycle can be incorporated into a digital twin forming a unified and accessible foundation for a wide range of offshore piles. Combined with monitored data, the proposed dynamic updated digital twin can be very important to provide rapid analysis results for reliable soil parameters and enables intelligent decision making on the pile behaviors.

Followed by Kapteyn et al. (2021), we can introduce *Rewards* and *Actions*. Partially observed probabilistic graphical model (POPGM) can be implemented in our own setting in Figure 4.1. This process can be roughly divided into two main parts: (1) calibration and assimilation; (2) prediction; as shown in Equation (4.1) and Equation (4.2), respectively:

$$\begin{aligned} &p(D_0, ..., D_{t_c}, Q_0, ..., Q_{t_c}, R_0, ..., R_{t_c} | o_0, ..., o_{t_c}, u_0, ..., u_{t_c}) \\ &= \prod_{t=0}^{t_c} [\phi_t^{update} \phi_t^{QoI} \phi_t^{evaluation}] \end{aligned} \tag{4.1}$$

$$\begin{aligned} &p(D_0, ..., D_{t_p}, Q_0, ..., Q_{t_p}, R_0, ..., R_{t_p}, U_{t_c+1}, ..., U_{t_p} | o_0, ..., o_{t_c}, u_0, ..., u_{t_c}) \\ &\propto \prod_{t=0}^{t_p} [\phi_t^{dynamics} \phi_t^{QoI} \phi_t^{evaluation}] \prod_{t=0}^{t_c} \phi_t^{assimilation} \prod_{t=t_c+1}^{t_p} \phi_t^{control} \end{aligned} \tag{4.2}$$

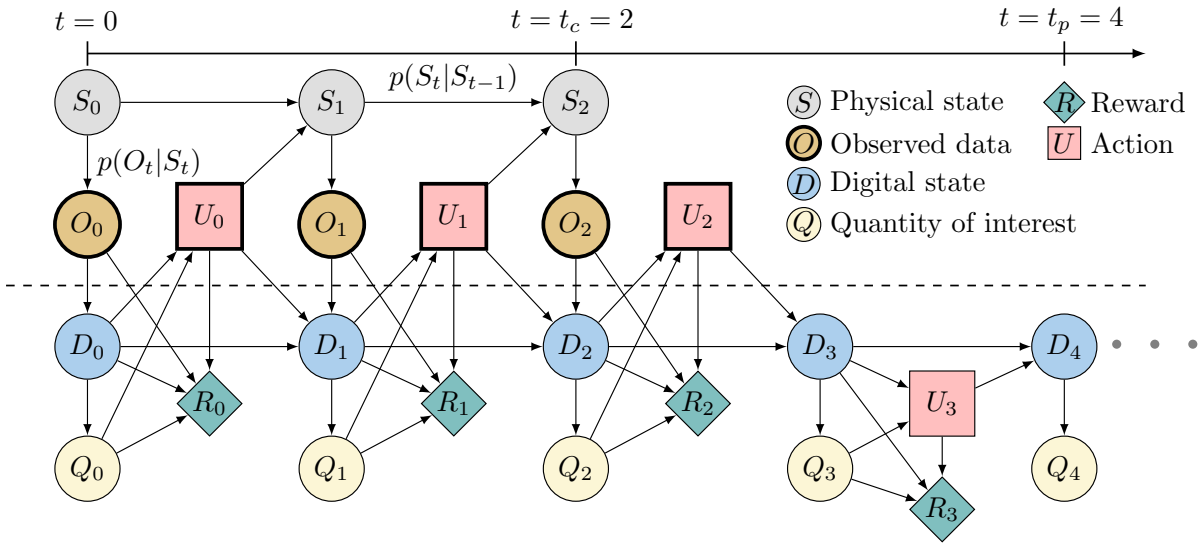


Figure 4.1: POPGM in digital twin

4.3 Next implementation for a pile:

Stage 1

Calibrate the models in Figure 4.2 with POPGM shown in Figure 4.1.

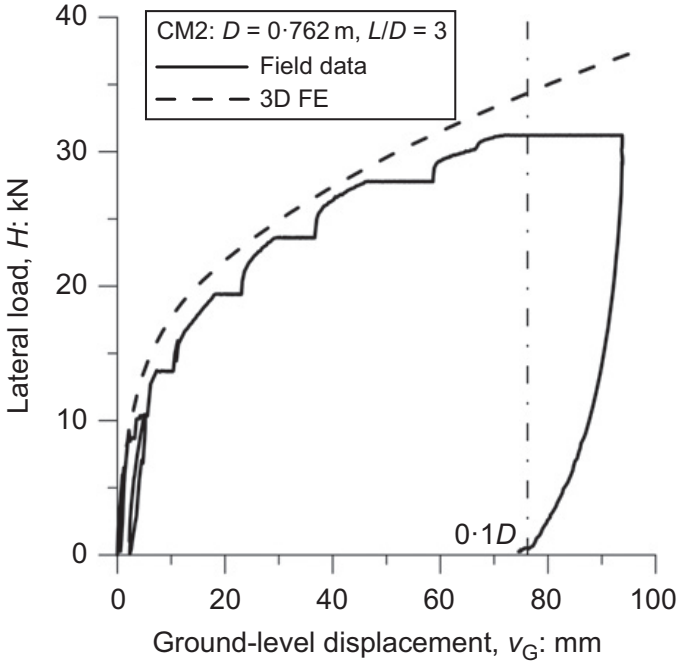


Figure 4.2: CM2 pile load displacement from Zdravković et al. (2020)

Computational forward model settings are listed:

- Software: Imperial College Finite Element Program (details can be seen in the link).

- Constitutive model: clay model adapted from Zdravković et al. (2020) and sand model adapted from Taborda et al. (2020).

Objective: Ensure the soil parameters in digital model can reveal unique characteristics of piles.

Stage 2

In operational phase, based on POPGM in Figure 4.1, continue the assimilation process: extend the digital twin capability to capture the piles response during loading.

Stage 3

Extension to prediction on the predictive pile performance.

4.4 Time plan

Table 4.1: PhD timeline

month	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48
Literature review	✓	✓	✓														
Numerical modelling (Data collection)		✓	✓	✓	✓	✓	✓										
Statistics Methods learning		✓	✓	✓	✓	✓	✓	✓	✓								
Statistics analysis calibration			✓	✓	✓												
Statistics analysis assimilation						✓	✓	✓	✓	✓	✓						
Statistics analysis prediction												✓	✓	✓			
Thesis writing															✓	✓	✓
Journal/Conference								✓				✓					✓

References

Andrieu, C., De Freitas, N., Doucet, A., & Jordan, M. I. (2003). An introduction to mcmc for machine learning. *Machine learning*, *50*, 5–43.

API. (2011). *Geotechnical and foundation design considerations.* API Washington, DC.

Bhattacharya, S. (2019). *Design of foundations for offshore wind turbines.* John Wiley & Sons.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association*, *112*(518), 859–877.

Buckley, R., Chen, Y. M., Sheil, B., Suryasentana, S., Xu, D., Doherty, J., & Randolph, M. (2023). Bayesian optimization for cpt-based prediction of impact pile drivability. *Journal of Geotechnical and Geoenvironmental Engineering*, *149*(11), 04023100.

Byrne, B. W., & Houlsby, G. T. (2003). Foundations for offshore wind turbines. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, *361*(1813), 2909–2930.

DenisBoigelot. (2011). *File:correlation examples2.svg.* Retrieved 2011-05-09, from https://commons.wikimedia.org/wiki/File:Correlation_examples2.svg?uselang=zh-cn

Finno, R. J., & Calvello, M. (2005). Supported excavations: observational method and inverse modeling. *Journal of geotechnical and geoenvironmental engineering*, *131*(7), 826–836.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.

Goodman, J., & Weare, J. (2010). Ensemble samplers with affine invariance. *Communications in applied mathematics and computational science*, *5*(1), 65–80.

Hsein Juang, C., Luo, Z., Atamturktur, S., & Huang, H. (2013). Bayesian updating of soil parameters for braced excavations using field observations. *Journal of Geotechnical and Geoenvironmental Engineering*, *139*(3), 395–406.

Jin, Y., Biscontin, G., & Gardoni, P. (2021). Adaptive prediction of wall movement during excavation using bayesian inference. *Computers and Geotechnics*, *137*, 104249.

Kapteyn, M. G., Pretorius, J. V., & Willcox, K. E. (2021). A probabilistic graphical model foundation for enabling predictive digital twins at scale. *Nature Computational Science*, *1*(5), 337–347.

Lataniotis, C. (2019). *Data-driven uncertainty quantification for high-dimensional engineering problems* (Unpublished doctoral dissertation). ETH Zurich.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Nakamura, K., Yamamoto, S., & Honda, M. (2011). Sequential data assimilation in geotechnical engineering and its application to seepage analysis. In *14th international conference on information fusion* (pp. 1–6).

Nguyen, L. T., & Nestorović, T. (2016). Nonlinear kalman filters for model calibration of soil parameters for geomechanical modeling in mechanized tunneling. *Journal of Computing in Civil Engineering*, *30*(2), 04015025.

Randolph, M., Cassidy, M., Gourvenec, S., & Erbrich, C. (2005). Challenges of offshore geotechnical engineering. In *Proceedings of the international conference on soil mechanics and geotechnical engineering* (Vol. 16, p. 123).

Randolph, M., & Gourvenec, S. (2017). *Offshore geotechnical engineering*. CRC press.

Royston, R., Sheil, B. B., & Byrne, B. W. (2022). Undrained bearing capacity of the cutting face for an open caisson. *Géotechnique*, *72*(7), 632–641.

Speich, M., Dormann, C. F., & Hartig, F. (2021). Sequential monte-carlo algorithms for bayesian model calibration–a review and method comparison. *Ecological Modelling*, *455*, 109608.

Stuyts, B., Weijtjens, W., & Devriendt, C. (2023). Development of a semi-structured database for back-analysis of the foundation stiffness of offshore wind monopiles. *Acta Geotechnica*, *18*(1), 379–393.

Sudret, B. (2019). Polynomial chaos expansions. *4th National Conference on Multidisciplinary Design, Analysis and Optimization*.

Taborda, D. M., Zdravković, L., Potts, D. M., Burd, H. J., Byrne, B. W., Gavin, K. G., ... others (2020). Finite-element modelling of laterally loaded piles in a dense marine sand at dunkirk. *Géotechnique*, *70*(11), 1014–1029.

Tang, C., Cao, Z.-J., Hong, Y., & Li, W. (2023). State space model of undrained triaxial test data for bayesian identification of constitutive model parameters. *Géotechnique*, 1–15.

Tao, Y.-q., Sun, H.-l., & Cai, Y.-q. (2021). Bayesian inference of spatially varying parameters in soil constitutive models by using deformation observation data. *International Journal for Numerical and Analytical Methods in Geomechanics*, *45*(11), 1647–1663.

Torre, E., Marelli, S., Embrechts, P., & Sudret, B. (2019). Data-driven polynomial chaos expansion for machine learning regression. *Journal of Computational Physics*, *388*, 601–623.

Verleysen, M., & François, D. (2005). The curse of dimensionality in data mining and time series prediction. In *International work-conference on artificial neural networks* (pp. 758–770).

Wagner, P.-R., Fahrni, R., Klippel, M., Frangi, A., & Sudret, B. (2020). Bayesian calibration and sensitivity analysis of heat transfer models for fire insulation panels. *Engineering structures*, *205*, 110063.

Wagner, P.-R., Nagel, J., Marelli, S., & Sudret, B. (2022). *UQLab user manual – Bayesian inversion for model calibration and validation* (Tech. Rep.). Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland. (Report UQLab-V2.0-113)

Wang, M., Wang, C., Hnydiuk-Stefan, A., Feng, S., Atilla, I., & Li, Z. (2021). Recent progress on reliability analysis of offshore wind turbine support structures considering digital twin solutions. *Ocean Engineering*, *232*, 109168.

Zdravković, L., Jardine, R. J., Taborda, D. M., Abadias, D., Burd, H. J., Byrne, B. W., ... others (2020). Ground characterisation for pisa pile testing and analysis. *Géotechnique*, *70*(11), 945–960.

Zhao, X., Dao, M. H., & Le, Q. T. (2023). Digital twining of an offshore wind turbine on a monopile using reduced-order modelling approach. *Renewable Energy*, *206*, 531–551.