

FEDERAL UNIVERSITY OF JUIZ DE FORA  
EXACT SCIENCES INSTITUTE  
BACHELOR'S DEGREE IN COMPUTER SCIENCE

# **An Improved Approach for Uncertainty Quantification in Enhanced Oil Recovery**

**Gabriel Brandão de Miranda**

JUIZ DE FORA  
AUGUST, 2022

# An Improved Approach for Uncertainty Quantification in Enhanced Oil Recovery

GABRIEL BRANDÃO DE MIRANDA

Federal University of Juiz de Fora  
Exact Sciences Institute  
Department of Computer Science  
Bachelor's Degree in Computer Science

Supervisor: Bernardo Martins Rocha  
Co-supervisor: Rodrigo Weber dos Santos

JUIZ DE FORA  
AUGUST, 2022

# AN IMPROVED APPROACH FOR UNCERTAINTY QUANTIFICATION IN ENHANCED OIL RECOVERY

Gabriel Brandão de Miranda

MONOGRAPHY SUBMITTED TO THE FACULTY OF THE EXACT SCIENCES INSTITUTE OF FEDERAL UNIVERSITY OF JUIZ DE FORA, AS AN INTEGRAL PART OF THE REQUIREMENTS NECESSARY TO OBTAIN THE BACHELOR'S DEGREE IN COMPUTER SCIENCE.

Approved by:

Bernardo Martins Rocha  
D.Sc. in Computational Modeling

Rodrigo Weber dos Santos  
D.Sc. in Mathematics

Grigori Chapiro  
D.Sc. in Mathematics

Iury Higor Aguiar da Igreja  
D.Sc. in Computational Modeling

JUIZ DE FORA  
2022, AUGUST 08

# Abstract

Knowledge about the uncertainties associated with a computational model is essential to understand the level of reliability of a prediction, especially in the context of decision making. The model's reliability is a central object in the scope of uncertainty quantification (UQ) techniques, which focus on discussing the uncertainty of predictions and guaranteeing more reliability to the simulations.

In the petroleum engineering field it is no different, the need for reliable simulations is mandatory, especially as the drilling and extraction of new wells becomes more expensive. In the context of enhanced oil recovery (EOR) processes, the foam injection technique is used to reduce gas mobility and increase the apparent viscosity, which in turn increases recovery efficiency. Uncertainty from input parameters may affect the simulators' reliability, and therefore uncertainty quantification is a key procedure while studying these physical systems.

This work presents an improved workflow for UQ in mathematical models used for the simulation of EOR processes. A study case based on experimental data from core flooding experiment is investigated with the aim of estimating the input parameters and their uncertainty. In particular, the CMG-STARS implicit texture foam model is adopted considering the term that describes the presence of water saturation in the foam. Different strategies for the Bayesian modeling and estimation are studied in order to increase the level of information available propagated to the whole process of UQ.

**Keywords:** Uncertainty quantification, Computational modeling, Bayesian modeling, Enhanced oil recovery.

## Resumo

O conhecimento sobre as incertezas associadas a um modelo computacional é essencial para entender o nível de confiabilidade de suas previsões, principalmente para tomada de decisão. A confiabilidade do modelo é o foco nas técnicas de quantificação de incerteza (UQ), que focam em discutir a incerteza das previsões e garantir mais confiabilidade às simulações.

Na área de engenharia de petróleo essa questão não é diferente, a necessidade de simulações confiáveis é obrigatória, principalmente porque a perfuração e extração de novos poços se torna cada vez mais cara. No contexto dos processos de recuperação avançada de óleo (EOR), a técnica de injeção de espuma é usada para reduzir a mobilidade do gás e aumentar a viscosidade aparente, o que, por sua vez, aumenta a eficiência da recuperação. A incerteza dos parâmetros de entrada pode afetar a confiabilidade dos simuladores e, portanto, a quantificação da incerteza é um procedimento fundamental ao estudar esses sistemas físicos.

Este trabalho apresenta um aprimoramento para UQ em modelos matemáticos usados na simulação de processos de recuperação avançada de petróleo. Um caso de estudo baseado em dados experimentais do experimento de inundação de testemunhos é investigado com o objetivo de estimar os parâmetros de entrada e suas incertezas. Em particular, o modelo de espuma de textura implícita CMG-STARS é adotado considerando o termo que descreve a presença de saturação de água na espuma. Diferentes estratégias de modelagem e estimação bayesiana são estudadas a fim de aumentar o nível de informação disponível propagada para todo o processo de UQ.

**Palavras-chave:** Quantificação de incertezas, Modelagem computacional, Modelagem Bayesiana, Recuperação avançada de petróleo.

# Thanks

To everyone who somehow contributed to my formation, not only as a researcher, but as an individual.

**Acknowledgments:** This research was carried out in association with the ongoing R&D project registered as ANP 20715-9, "Modelagem matemática e computacional de injeção de espuma usada em recuperação avançada de petróleo", Universidade Federal de Juiz de Fora (UFJF)/Shell Brasil / ANP. Mathematical and computational modeling of foam injection as an enhanced oil recovery technique applied to Brazil pre-salt reservoirs, sponsored by Shell Brasil under the ANP R&D levy as "Compromisso de Investimentos com Pesquisa e Desenvolvimento". This project is carried out in partnership with Petrobras.

*“The ears hearing something is not as good as the eyes seeing it;*

*The eyes seeing it is not as good as the feet treading upon it;*

*The feet treading upon it is not as good as the hands differentiating it.*

*When a man first enters an office, it is as if he has entered a dark room: the longer he stays, the more clearly he sees.”*

ATTRIBUTED TO *Xunzi*, in the *Shuo Yuan*.

# Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Uncertainty Quantification . . . . .	7
1.2	Petroleum Engineering . . . . .	8
1.3	Objectives . . . . .	9
1.4	Organization of the text . . . . .	10
<b>2</b>	<b>Uncertainty Quantification Workflow</b>	<b>11</b>
2.1	The Workflow . . . . .	11
2.2	Probabilistic Modeling . . . . .	12
2.2.1	A Brief Review on Probability Theory . . . . .	12
2.2.2	Bayesian Approach . . . . .	13
2.2.3	Hierarchical Modeling . . . . .	14
2.2.4	Probabilistic Programming . . . . .	15
2.3	Bayesian Inference . . . . .	16
2.3.1	Inference Engines . . . . .	17
2.4	Diagnostics and Evaluation . . . . .	19
2.5	Sampling Methods . . . . .	22
2.5.1	Monte Carlo Methods . . . . .	22
2.5.2	Emulators . . . . .	24
2.6	Sensitivity Analysis . . . . .	24
2.6.1	Sobol Indices . . . . .	25
2.7	An illustrative example . . . . .	25
2.7.1	Inverse Uncertainty Quantification . . . . .	27
2.7.2	Forward Uncertainty Quantification . . . . .	28
2.7.3	Sensitivity Analysis . . . . .	29
2.7.4	Concluding Remarks . . . . .	30
<b>3</b>	<b>Foam Assisted Enhanced Oil Recovery</b>	<b>31</b>
3.1	Background . . . . .	31
3.2	Mathematical Modeling . . . . .	32
3.2.1	Fractional Flow Theory . . . . .	32
3.2.2	Relative Permeabilities . . . . .	33
3.2.3	Foam Modeling . . . . .	33
3.2.4	Experimental Data . . . . .	34
3.2.5	Hierarchical Modeling and Noninformative Priors . . . . .	35
3.2.6	On the choice of prior distributions for variance parameter . . . . .	37
3.3	Results . . . . .	40
3.3.1	Experiment I - Maximum Variance . . . . .	40
3.3.2	Experiment II - Maximum Likelihood Estimation (MLE) of Variance . . . . .	41
3.3.3	Experiment III - Weighting the Variance . . . . .	41
3.3.4	Experiments Comparison . . . . .	43
3.3.5	Sensitivity Analysis . . . . .	44
3.4	Experiment Conclusion . . . . .	44

<b>4 Conclusions</b>	<b>47</b>
4.1 Limitations and Future Works . . . . .	47
<b>Bibliography</b>	<b>49</b>

# 1 Introduction

This chapter introduces basic concepts of uncertainty quantification, reservoir simulation in the context of petroleum engineering and the objectives of this work.

## 1.1 Uncertainty Quantification

Uncertainty in science can be characterized as the lack of certainty about a phenomenon. More precisely, it is a notion of limitation of knowledge about some event because of its physical nature. For the output of a model, it is represented by a confidence level with respect to the model's predictions. Uncertainties can be classified into two types:

- *Aleatoric*: which are uncertainties due to variations in the physical system itself. They arise naturally from observations (noise), they are stochastic and irreducible;
- *Epistemic*: which uncertainties due to the lack of knowledge of a physical system. This form of uncertainty arises from modeling hypothesis (bias) and can be reduced.

Uncertainty Quantification (UQ) is the science of quantitative characterization and reduction of uncertainties in both computational and real world applications. To deal with massive datasets in diverse application areas, data-driven uncertainty quantification requires statistical tools such as Monte Carlo methods, Bayesian inference, and Markov decision processes.

UQ can be further categorized into Inverse and Forward UQ. The former is the process of estimating the parameters of a model and their uncertainty from experimental data. Inverse problems are usually ill-posed, and their solution not unique. Many applications employ inverse UQ for calibrating models of physical systems, such as computational fluid dynamics (CFD) (LI et al., 2022; VALDEZ et al., 2020) and computational structural dynamics (CSD) (ALVIN et al., 1998; BALLESTEROS et al., 2014; SEDEHI; PAPADIMITRIOU; KATAFYGIOTIS, 2020) simulations. Forward UQ estimates the uncertainty in the model's output given the uncertainty in the inputs, which are defined as random vari-

ables. A statistical analysis of the outputs can be carried out to quantify the propagated uncertainty.

Advances in UQ algorithms in recent years have made it feasible to infer parameters and their uncertainties in large-scale complex systems from large datasets, such as global ocean modeling and oil recovery modeling (COUNCIL et al., 2012; RUDE et al., 2018; BERG; UNSAL; DIJK, 2021).

Simulations have transformed the petroleum sector, where a wrong drilling decision can take up the entire available budget for a project. The flow of fluids between injection and production wells can be studied using reservoir simulation tools. Simulating numerous forward scenarios is necessary to optimize reservoir exploitation while lowering uncertainty. In fact, the initiatives to enhance reservoir simulators have been funded by oil companies, which are responsible for some of the most powerful computers (RUDE et al., 2018).

## 1.2 Petroleum Engineering

Oil recovery is one of the most critical applications in studying fluid flow in porous media. Finding more effective ways to extract oil from natural reservoirs is a major concern for the petroleum sector. The oil recovery process can be seen divided in three main types:

- *Primary recovery*: Corresponds to the natural oil flow from the reservoir to the surface. It is the first and most crucial step in the oil recovery process. Generally, the amount of oil recovered by primary recovery is about 15% to 30% of the original oil on site (OOIP).
- *Secondary recovery*: More oil is recovered by injecting water or gas into the reservoir so that the pressure remains high enough to maintain oil extraction in a production well. Despite increasing efficiency, this strategy still suffers from instability in the gas-water interface and gravity segregation that create viscous fingering.
- *Tertiary recovery*: After the secondary recovery is performed, more oil can be recovered by more efficient strategies, such as chemical or thermal injections.

It is important to remark that Enhanced Oil Recovery (EOR) is a process that can be used to increase the amount of oil recovered since the primary injection employing secondary and tertiary techniques.

This work focus in a particular case about the interaction between aqueous and gaseous phases that creates foam, a dispersion of gas in a liquid stabilized by surfactants. Foam is a very efficient method for oil recovery, because it can reduce the mobility of the gas phase and increase the sweep efficiency of the reservoir (EFTEKHARI; FARA-JZADEH, 2017). Figure 1.1 presents an illustrative scheme comparing two techniques: gas injection and the co-injection of water and gas generating foam.

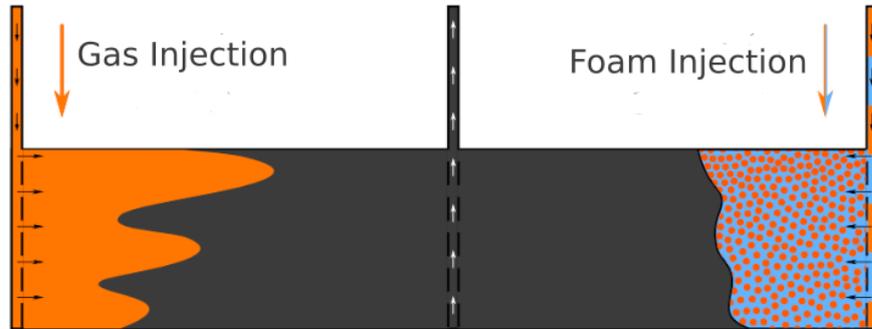


Figure 1.1: A comparison of gas injection and foam injection as examples of secondary and tertiary or EOR. Adapted from (CEDRO, 2020).

## 1.3 Objectives

The main objective of this work is to present the usage of UQ as an essential tool for estimation of parameters and uncertainties in a computational model for foam flow in EOR processes. In particular, the parameters of the mathematical model known as STARS ((CMG)., 2019) and proposed by the Canadian company Computer Modelling Group (CMG) are estimated using the UQ workflow presented in this work. An evaluation of how to improve the parametric uncertainty estimation using different Bayesian modeling choices is also carried out.

---

## 1.4 Organization of the text

Chapter 2 presents a detailed overview of the workflow for performing UQ in computational models, separated into stages to facilitate the understanding of and the role of each stage. At the end of Chapter 2, an illustrative example for carrying out UQ is presented to highlight the steps of the UQ workflow. Chapter 3 presents a case study of UQ to estimate parameters and reduce their uncertainty considering experimental data from a foam quality-scan experiment. General conclusions about the present work and possible future works are discussed in Chapter 4.

## 2 Uncertainty Quantification Workflow

This chapter presents a detailed overview of the usage of UQ on mathematical models. Some necessary concepts and theories are given to introduce the methods employed in this work. First of all, to have an overview of the steps that make up the UQ pipeline, it is essential to know the primary purpose of employing the UQ on a model: to understand how the system modeled reacts to uncertain inputs.

### 2.1 The Workflow

A workflow for uncertainty quantification is presented in Figure 2.1. Equations, drawings, and symbols in Figure 2.1 are associated with the context of multi-phase flows in porous media since it is the focus of this work.

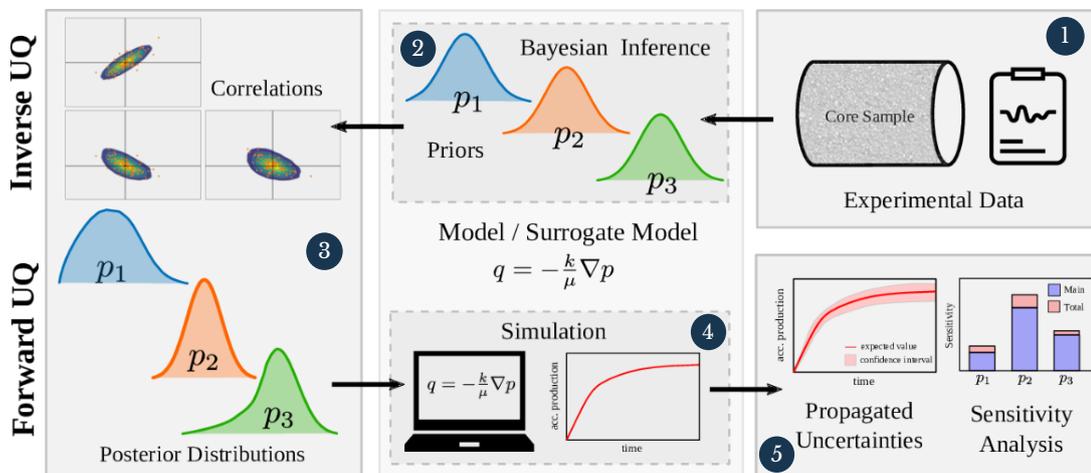


Figure 2.1: Workflow for uncertainty quantification of numerical models for foam-based EOR presented in Rio Oil and Gas 2022. Adapted from (RIBEIRO et al., 2022).

In the following it is described in well-defined steps the different techniques employed in the workflow:

1. **Probabilistic Modeling** the uncertain data, the input parameters and its corre-

- lations, and also the prior knowledge available for the researcher;
2. **Bayesian Inference** over the evidence (data) and prior beliefs;
  3. **Diagnostics and Evaluation** of the posteriors distributions of parameters values;
  4. **Uncertainty Propagation** of posteriors distributions to the model;
  5. **Sensitivity Analysis** of the output led by changes in input variables.

The task of performing inverse UQ analysis is characterized by steps 1, 2, and 3, whereas forward UQ analysis is characterized by steps 3, 4, and 5 from the workflow. In the following sections, each of these steps are discussed with more details. At the end of this chapter, an illustrative example of the application of this workflow to a classical mathematical model problem is shown to highlight the presented concepts.

## 2.2 Probabilistic Modeling

Probabilistic modeling characterizes the variability of model parameters (inputs) or even mathematical expressions that compose it in terms of random variables. It occurs through randomization processes, employing well-known probability distributions for the parameters or components of the model.

In this section, before presenting how the mathematical model randomization is done, some concepts of probability theory are briefly reviewed. A more in-depth and detailed explanation can be found on the literature (CASELLA; BERGER, 2021).

### 2.2.1 A Brief Review on Probability Theory

A probability  $\mathbb{P}(X = x) \in [0, 1]$  defines the chance of an event  $x$  occurring. Here  $X$  is a random variable, a function that maps a probability space to the real numbers. Random variables are associated with probability distributions. Probability distributions are mathematical functions that define the probabilities of occurrence of a variable taking any given value. In the continuous cases, it is defined as probability density functions (PDF). In the literature, there are many well-known probability distribution families, such as the Gaussian or normal distribution, detailed next. Considering  $X$  follow a

normal distribution, a common notation used to indicate this fact and to describe its PDF is:

$$X \sim \mathcal{N}(\mu, \sigma^2) \quad \iff \quad f_X(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

where  $\mu$  and  $\sigma^2$  are the parameters of the normal distribution, which are also statistics of the distribution of population. The parameter  $\mu$  represents its mean (or expectation), while  $\sigma^2$  denotes its variance. Statistics are summaries of populations or samples of those populations, and therefore, being summaries, they are simpler objects to deal with.

Two random variables  $X_1$  and  $X_2$  are said to be independent if the realization of one does not affect the probability distribution of the other. If they are not independent, the probability can be measured by the conditional probability, i.e., the probability of an event  $X_1$  is dependent on the occurrence of another event  $X_2$ . Formally defined by  $\mathbb{P}(X_1|X_2) = \frac{\mathbb{P}(X_1 \cap X_2)}{\mathbb{P}(X_2)}$ .

If  $X_1$  and  $X_2$  are independent, the correlation coefficient represents a measure of the similarity between these two variables is denoted by  $Corr(X_1, X_2)$  and is a number  $[-1, 1]$ . A positive correlation between two random variables can be interpreted as that as one variable increases, the other increases. In contrast, with a negative correlation when one variable increases, the other decreases.

Finally, probabilistic modeling is a branch of the probability and statistics that deals with developing mathematical models for analyzing data, generally subject to uncertainty. It is also about dealing with the randomization of constituent parameters of a deterministic model, i.e., characterizing variables as random variables.

### 2.2.2 Bayesian Approach

The probabilistic modeling of a physical system with observed data can be carried out using the Bayesian approach, glimpsed as a **Bayesian modeling**, which is a powerful tool for analyzing complex data. The Bayesian modeling is a statistical approach that uses the Bayes theorem as its foundation to answer how this data could have been generated, given some data and assumptions.

The Bayesian model is composed of combining random variables and the data information via the Bayes' theorem:

$$\underbrace{\mathbb{P}(\theta|D)}_{\text{posterior}} = \frac{\overbrace{\mathbb{P}(D|\theta)}^{\text{likelihood}} \overbrace{\mathbb{P}(\theta)}^{\text{prior}}}{\underbrace{\mathbb{P}(D)}_{\text{marginal likelihood}}}, \quad (2.1)$$

where  $\mathbb{P}(\theta)$  represents the prior knowledge of the input parameters  $\theta$ ;  $\mathbb{P}(D|\theta)$  is the likelihood function is how the observed data  $D$  is introduced on model; and  $\mathbb{P}(D)$  is the marginal likelihood or evidence, a normalization factor for the posterior distribution,  $\mathbb{P}(\theta|D)$ . This equation will be revisited in more detail soon.

In Bayesian theory, the probability reflects the degree of confidence in an event, which may change when new information is collected, as opposed to a fixed value based on a frequentist approach. The degree of confidence can be established on prior knowledge about the event, such as results from previous experiments or personal belief in the event.

### 2.2.3 Hierarchical Modeling

Bayesian hierarchical modeling is a central topic in modern Bayesian statistics. Its construction occurs by combining submodels into a hierarchical structure and employing Bayes' theorem to associate them with the observed data, considering any uncertainties present (MARTIN; KUMAR; LAO, 2021; MCELREATH, 2020).

The validation of a simulator or a model denoted by  $\mathcal{M}$  via the Bayesian approach is based on the assumption that the error of the data against the model,  $\epsilon$ , follows the normal distribution centered at origin with a given variance, i.e.,  $\epsilon \sim \mathcal{N}(0, \sigma^2)$ . Applying that error at the model's output,  $\beta$ , then the random variable representing the model is

$$\mathcal{M} = \epsilon + \beta \implies \mathcal{M} \sim \mathcal{N}(\mu, \sigma^2). \quad (2.2)$$

Equation (2.2) presents a Bayesian regression where the regressand is the investigated model  $\beta$ . In this way, when randomizing the set of parameters of the model, its outputs,  $\beta$ , will become another random variable because it inherits the randomness when operating over them. Figure 2.2 illustrates the hierarchical composition for a case with two

randomized parameters

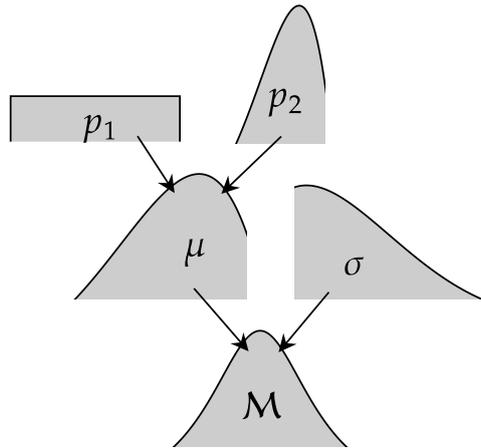


Figure 2.2: A simplified Kruschke style diagram (KRUSCHKE, 2015) exposing the relations on a idealized hierarchical modeling that composes model  $\mathcal{M}$ . To exemplify, in the illustrated case, for a hypothesis that  $\mathcal{M}$  follows a normal distribution. Parameters  $p_1$  and  $p_2$  are any random variables following a prior PDF that drive randomness on mean  $\mu$ . Also, the standard deviation  $\sigma$  has no influence from other random variables, so a prior PDF must be assumed.

To support Bayesian hierarchical modeling when dealing with complex systems, many computational implementations have emerged in recent years with sophisticated approaches, allowing practitioners to work with this approach focusing on the modeling only (SALVATIER; WIECKI; FONNESBECK, 2016).

### 2.2.4 Probabilistic Programming

Modern Bayesian statistics makes use of growing computational power to enable iterative algorithms that were unfeasible decades ago. Thus, with the growth of the use of computers, computational tools that aim to facilitate these constructions emerged under the umbrella of the name probabilistic programming languages (PPL). The notion of probabilistic programming is in the sense that probability models are created using programming random variables as the model's components. According to (MARTIN; KUMAR; LAO, 2021) “*the purpose of PPLs is help Bayesian practitioners build generative models to solve problem at hand, for example, perform inference of a Bayesian model*

through estimating the posterior distribution”. In this work the PPL adopted is the PyMC (SALVATIER; WIECKI; FONNESBECK, 2016), a Python package for Bayesian inference that employs Markov chain Monte Carlo (MCMC) methods.

## 2.3 Bayesian Inference

As presented in the previous section, Bayesian inference is about updating beliefs to incorporate new evidences (GELMAN et al., 2021) using the Bayes’ theorem (2.1), as illustrated in Figure 2.3. Therefore, it is essential to understand better how to calculate each Bayes theorems part to put it in computer language.

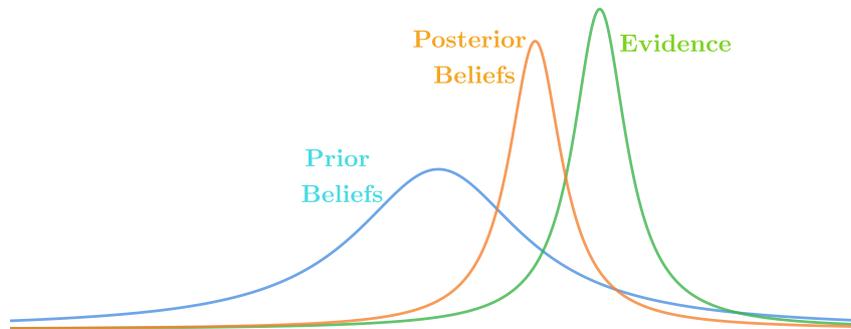


Figure 2.3: Probabilistic update from a prior beliefs to a posterior beliefs when confronted with evidences.

The prior distribution,  $\mathbb{P}(\theta)$ , translates the notion one have on that event  $\theta$ , in a sense that it represents the uncertainty about the parameters before observing the data. If nothing is known about that event, a flat prior could be used (later, it will be discussed that this is not necessarily a good choice).

To link the observed data  $D$  to the unknown parameters  $\theta$ , the likelihood function,  $\mathbb{P}(D|\theta) = \mathcal{L}(\theta | D)$  measures plausibility of data given the parameters. For a continuous case with an observed dataset  $D$ , the likelihood is the product of the PDF parameterized by  $\theta$  for all data  $d_i \in D$ :

$$\mathcal{L}(\theta | D) = \prod_{i=1}^n f_{\theta}(d_i), \quad (2.3)$$

where  $f_{\theta}(d_i)$  is the PDF of  $\theta$  evaluate in  $d_i$ .

As  $D$  is the observed data, formally  $\mathbb{P}(D)$  is the probability of observing the data in all universe of possibilities, for all the possible values the parameter  $\theta$  can take. In mathematical terms the marginal likelihood is

$$\mathbb{P}(D) = \int_{\Theta} \mathbb{P}(D | \theta) \mathbb{P}(\theta) d\theta, \quad (2.4)$$

where  $\Theta$  means an integration over all the possible values of  $\theta$ . This expression is hard to be calculated even with numeric approaches. But notice as it is a definite integral on the parameter space,  $\mathbb{P}(D)$  is constant for the parameters and, as pointed before, only a normalization factor for the posterior distribution.

So, in practice the Bayes' theorem is presented using a proportionality relation

$$\underbrace{\mathbb{P}(\theta|D)}_{\text{posterior}} \propto \overbrace{\mathbb{P}(D|\theta)}^{\text{likelihood}} \overbrace{\mathbb{P}(\theta)}^{\text{prior}} \quad (2.5)$$

By multiplying the prior distribution and likelihood, the posterior distribution is obtained to less than a normalization; the result of the Bayesian inference reflects the knowledge achieved over the problem given the data and the modeling.<sup>1</sup> In practice, when working with datasets, this approximation is the more feasible way to construct the Bayesian model.

### 2.3.1 Inference Engines

As seen so far, computing Bayesian methods can be mathematically challenging while conceptually simple. The posterior is usually estimated numerically using iterative approaches such as the Markov Chain Monte Carlo (MCMC), well-known in literature as an introductory algorithm for the universe of inference engines. An inference engine (also called Universal Inference Engines) is so named because it is designed to be capable of approximating any probabilistic model. (MARTIN; WIECKI, 2018) lists some of these methods separated into two convenient groups:

---

<sup>1</sup>The Bayesian uncertainty formulation makes no distinction between epistemic and aleatoric uncertainties. Instead, this distinction is interpretive and can be analyzed more carefully by methods that employ the result of this step.

- Non-Markovian methods:
  - Grid computing,
  - Quadratic approximation,
  - Variational methods;
- Markovian methods:
  - Metropolis-Hastings,
  - Hamiltonian Monte Carlo (HMC),
  - Sequential Monte Carlo (SMC).

More details about these inference engines and others can be found on the literature (MARTIN; WIECKI, 2018).

When a PPL is adopted, the possibilities of methods to be employed are reduced to the group inference engines a PPL provides. In this work we used the PyMC engine for Bayesian inference, which employs the No-U-Turn Sampler (NUTS) (HOFFMAN; GELMAN et al., 2014) based on the HMC (DUANE et al., 1987; BROOKS et al., 2011), a type of MCMC method that makes use of gradients to generate new proposed states.

To cover the core concepts of Markovian methods, the idea behind the classic MCMC sampling is discussed: The main goal of MCMC is to obtain a sufficient size of sample from a target distribution (in this case, the posterior distribution), to capture the essence of the true population distribution as the sample size increases. Thus, MCMC is a stochastic process that repeatedly draws random samples seeking to characterize the distribution of the parameter of interest. The Monte Carlo (MC) random walk method - to be discussed later - creates a Markov chain to have a balanced distribution proportional to a given function. This combination of methods, allows building a Markov chain that spans Monte Carlo samples, where probability dependencies between samples are taken into account.

In order to exemplify the execution of the method, listed below and illustrated in Figure 2.4 is the core of the execution of the first MCMC, the Metropolis algorithm (METROPOLIS et al., 1953; HASTINGS, 1970)

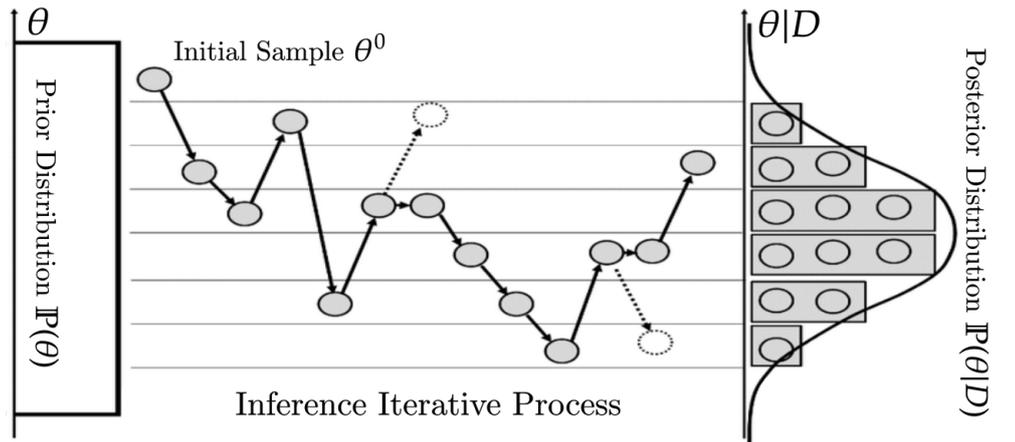


Figure 2.4: Example of the MCMC iterative process considering a flat prior. Adapted from (LEE; SUNG; CHOI, 2015)

1. Choose a starting point  $\theta^0$  of, usually arbitrarily;
2. Generates a candidate  $\theta^i$ ;
3. Sample a new parameter  $\theta^{i+1}$  sampling from a simpler distribution  $q(\theta^{i+1}|\theta^i)$ ;
4. Accepts or rejects this new point;
5. Repeat from step 2 until sufficient sample size is reached.

In general, more sophisticated MCMC-based strategies focus on the process of improving walking through population space, described in steps 2, 3 and 4. The term "sufficient" in step 5 refers to the quality of the model, this quality can be evaluated diagnosing the posterior samples. In the following section it will be discussed.

## 2.4 Diagnostics and Evaluation

This section covers the analysis and treatment of inverse UQ outputs and how to make them available as inputs to carry out forward UQ. An important first step is diagnosing inference results to understand if the chain has converged. In practice these diagnoses usually occur through graphical analyses and empirical tests which in turn are facilitated by the Python package `ArviZ` (KUMAR et al., 2019) and its interface with `PyMC`.

A usual analysis is the autocorrelation of the chains. This analysis is common in the context of time series forecasting to assess how related a series is to itself. Briefly,

given a correlation between two series  $T$  and  $S$  denoted  $Corr(T, S)$ , the autocorrelation function (ACF),  $R(k) \in [0, 1]$  measures the correlation of a series/chain with itself within a lag of size  $k$ ,

$$R(k) = Corr(T, T_k). \quad (2.6)$$

If the series is autocorrelated, it indicates a possible non-convergence. To frame a better understanding of the several values assumed by the lag  $k$ , a choice is to use the ACF plot, Figure 2.5. It is expected that will be no autocorrelation for a chain in convergence as  $k$  increases.

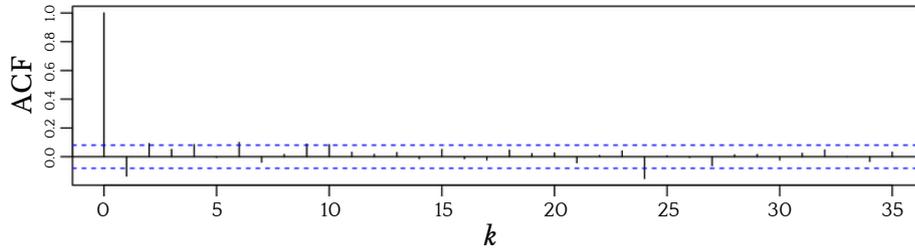


Figure 2.5: The Autocorrelation Function (ACF) plot for the trace of the distribution. Generally is defined an expected range  $R(k) < 0.1$  for all  $k > 1$  to evaluate the convergence.

Another useful analysis, that can also visually express the autocorrelation, is in the variable trace, if a behavior is observed where the points indicates oscillation of the moving average, can be an indication of non-convergence of the sample. In comparison the behavior will present a lean trace, which indicates a solid and constant moving average and consequently the convergence. These behavior of bad and good quality of the inference results by the trace plot are exemplified respectively in Figure 2.6.

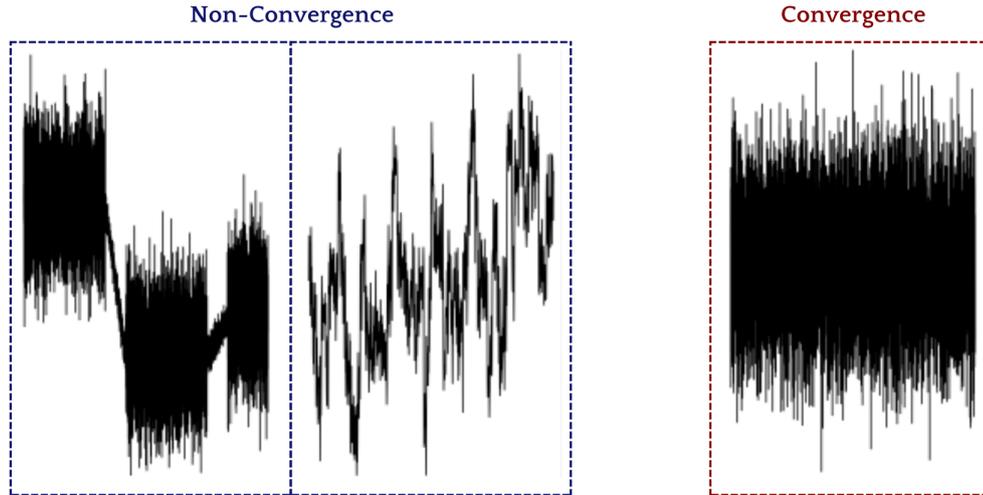


Figure 2.6: Trace plot examples that indicates bad quality, on the left, and good quality, on the right.

For now, one can already obtain relevant information about the model, mainly about the history matching on the parametrization with information about the uncertainty of this data assimilation. Furthermore, it is possible to observe how these parameters are correlated through the sampled chains, which provides an intuition about which parameters most influence the output uncertainty, or even how the proposed modeling for the phenomenon of interest can be reassessed. But this analysis will be saved for the end of the workflow.

To prepare the data to the forward UQ and avoid possible noise without interpretation when propagating inference results, the highest density interval (HDI) is employed. It is a way of summarizing a distribution by the shortest interval containing a fraction of a probability density (MARTIN; KUMAR; LAO, 2021). The HDI, seeks the region where is most likely to have an observation, but differs from confidence interval (CI) that removes the tails symmetrically in relation of probability. The Figure 2.7 illustrates these differences.

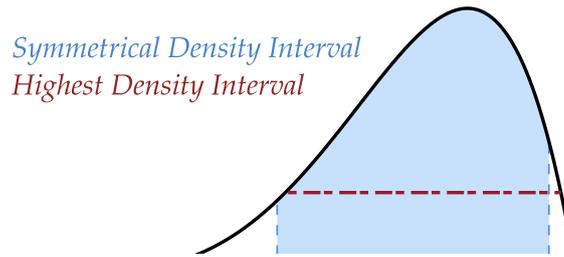


Figure 2.7: HDI has the same density value on its boundaries, different from a symmetrical interval.

Having the converging inference results diagnosed and curated to a meaningful credible interval, they are ready to be used as input artifacts for the forwarding UQ.

## 2.5 Sampling Methods

In this section will be discussed how to bring the uncertainties observed on the input parameters to the model output. As usual in the field of Statistics, this can be presented in a much more explanatory way through statistics, i.e., data reductions over a larger volume of information.

In order to obtain a probabilistic representation of the system's output, a large number of possible input parameters sets have to be analyzed. The uncertainty defined by the probability distribution of the input parameters simulates these scenarios when sampling from the input PDFs, systematically varying all the input parameters (scenario approach). The choice of the input parameters' probability distributions is problem-dependent and it is usually based on the expert's knowledge and data availability. One of first methods on random sampling is the Monte Carlo (MC) method (METROPOLIS; ULAM, 1949) that have inspired a class of other methods and will be presented in the following.

### 2.5.1 Monte Carlo Methods

Methods based on MC rely on the idea of random sampling to achieve a numerical result. In the context of UQ, MC methods can be employed to sample from posteriors

distributions and achieve a prediction interval (PI) for a quantity of interest (QoI). With sufficient independent samples from a distribution, it is possible to summarize the data with statistics such as mean and percentiles.

A classic example of MC usage is  $\pi$  constant estimation: Suppose a circle of diameter  $d$  enclosed by square of same side length. Knowing that the area of a circle is  $A_{\text{CIRCLE}} = \pi r^2$ , for this circle  $A_{\text{CIRCLE}} = \pi d^2/4$ . To approximate this value an identically distributed random variable over the range  $[0, d]$  - the uniform distribution - is independently sampled for the axes  $x$  and  $y$  enough times. For each point that presents a Euclidean distance from the center of circle smaller than the radius  $r = d/2$ , it is flagged that this point is inside the circle. At the end of the sampling, the proportion of points flagged is an estimate for

$$\frac{A_{\text{CIRCLE}}}{A_{\text{SQUARE}}} = \frac{\pi d^2/4}{d^2} = \frac{\pi}{4} \quad \implies \quad \pi = 4 \frac{A_{\text{CIRCLE}}}{A_{\text{SQUARE}}}.$$

Notice that the probability for a given point in  $[0, d] \times [0, d]$  being inside the circle is also  $\pi/4$ . Figure 2.8 illustrates this computation. The quality of this approximation depends on the number of samples.

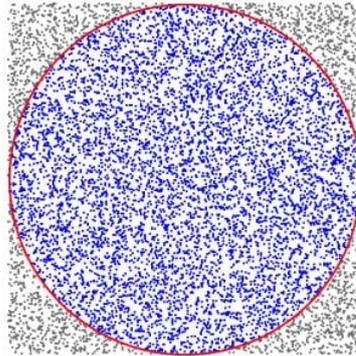


Figure 2.8: Random samples from a  $[0, d] \times [0, d]$  interval, where the observations identified within the circle are highlighted in blue.

MC method has a very simple computational implementation as it relies on the same idea of pseudo random numbers generation. So its usage in `Python` can be done via several famous `Python` packages, as for example `NumPy` (HARRIS et al., 2020) and `Pandas` (TEAM, 2020). As said before, there is much more to explore on this topic, where some

popular strategies are based on the idea of the MC method. To go further into this topic, see (LOHR, 2021).

In the same way, sampling on the inference results and propagating it to our QoI makes possible to evaluate statistics. A usual choice of reductions for sampled data is the median value, which can be extracted directly from the posterior distributions of parameters, that will be defined as best fitting parameters for the model. Also, to better understand the significance of the prediction, upper and lower boundaries defined via percentiles,  $[\ell, u]$ , can represent a prediction interval (PI) of a given confidence level  $\gamma$ , such as  $\mathbb{P}(\ell < X < u) = \gamma$ .

### 2.5.2 Emulators

In some cases it is very expensive evaluating some QoIs, strategies to make a smarter sampling from the inference results are some of the tools in the literature. Another common option is to use emulators, statistical models that seek to mimic the output of a complex physical system at a much lower computational cost (GRAMACY, 2020). Some examples of emulators are the polynomial chaos expansion (PCE) (XIU; KARNI-ADAKIS, 2002; GHANEM; SPANOS, 1991; WIENER, 1938) and the Gaussian process (GP) emulators (GRAMACY, 2020).

## 2.6 Sensitivity Analysis

Sensitivity analysis (SA) is the study of how the variation (uncertainty) in the output of a mathematical model, quantitatively or descriptively, is divided into different sources of variation in the model input. In more general terms, UQ and SA investigate the robustness of a study when the study includes some form of mathematical modeling. While UQ examines the complete uncertainty of the study's conclusions, SA attempts to determine which source of uncertainty weighs more in the study's conclusions. Sensitivity analysis in uncertainty quantification context is a powerful tool to assess the impact of the model parameters' uncertainty on the model output and achieve interpretability.

There are different strategies to analyze the sensitivity of a model, as for example

via its derivative or via its variance. In this work will be focused on the second one.

### 2.6.1 Sobol Indices

Sobol indices, or variance-based sensitivity analysis, are metrics from the family of statistical methods for analysis of variance (ANOVA). It is used to quantify the relative importance of the input parameters on the output via a decomposition of the model output's variance into fractions which can be attributed to inputs or sets of inputs (SOBOL', 1990; SOBOL, 2001; SALTELLI et al., 2008). More specifically, the first-order Sobol index (or main Sobol index) evaluates the influence of a given input parameter  $\theta_i \in \Theta$  independently in the output. The total Sobol index evaluates the direct contribution and interactions among parameters to the variance of the output  $\mathcal{Y}$ . The following expressions describe these Sobol indices respectively:

$$S_{1_i} = \frac{\mathbb{V}[\mathbb{E}[\mathcal{Y}|\theta_i]]}{\mathbb{V}[\mathcal{Y}]}, \quad \text{and} \quad S_{T_i} = 1 - \frac{\mathbb{V}[\mathbb{E}[\mathcal{Y}|\theta_{-i}]]}{\mathbb{V}[\mathcal{Y}]}, \quad (2.7)$$

where  $\mathbb{V}[\cdot]$  is the variance and  $\mathbb{E}[\cdot]$  the expected value.  $\theta_{-i}$  denotes the set of all input parameters except the  $\theta_i$  parameter.

## 2.7 An illustrative example

A simple mathematical model is considered here to illustrate how the UQ workflow can be used. The logistic growth model, given by the following ordinary differential equation (ODE),

$$\frac{dP}{dt} = rP \left( 1 - \frac{P}{K} \right), \quad t \in [0, T] \quad (2.8)$$

where  $P$  is the variable of interest that represents the size of a given population;  $t$  is the independent variable for the time;  $r$  is the growth rate of the population and  $K$  is the population's carrying capacity. For a given initial population  $P(0) = P_0$ , the ODE has an unique solution.

We considered a synthetica case to evaluate the UQ workflow for parameter inference. The ground truth parameter values are set to  $K = 10$ ,  $r = 0.1$ , and  $P_0 = 0.1$ .

Synthetic data are generated through an approximate solution over a discretized interval  $t \in [0, 120]$ . Then, noise is added using a normal distribution, where each point generated without noise,  $P_i^{\text{GEN}}$ , is computed as a random sample from

$$P_i^{\text{OBS}} \sim \mathcal{N}(P_i^{\text{GEN}}, 0.05P_i^{\text{GEN}}).$$

The generated (synthetic) data is presented is shown in Figure 2.9.

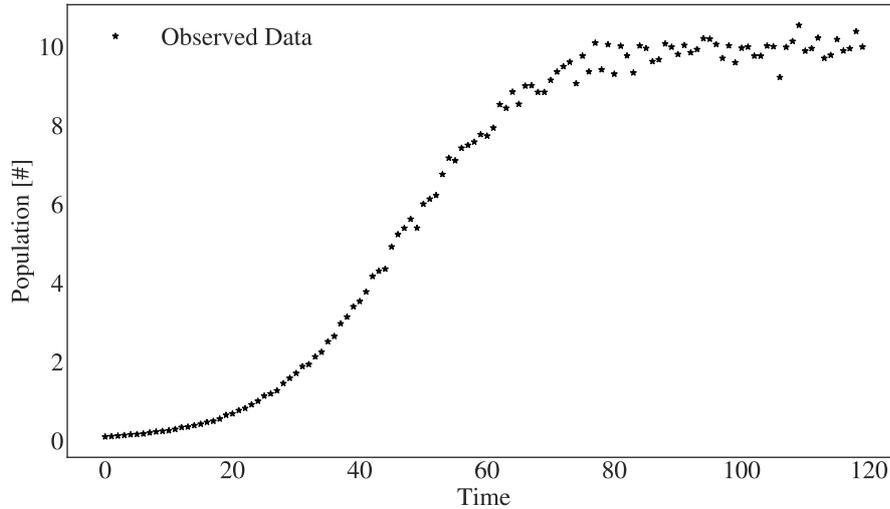


Figure 2.9: Synthetic data generated with ground truth parameters.

In this example,  $r$  and  $K$  are the parameters to be randomized. Seeking to make a conservative choice and keep the Bayesian modeling simple, a uniform distribution is adopted in a convenient interval for both parameters:

$$K \sim \mathcal{U}[0.8 \cdot P^K, 2 \cdot P^K] \quad \text{and} \quad r \sim \mathcal{U}[0.01, 10], \quad (2.9)$$

where,  $P^K = \max(P^{\text{OBS}})$ .

The Bayesian model is completed assuming the output follows a normal distribution with its scale parameter referring to the error on the simulator,

$$P \sim \mathcal{N}(P^{\text{DM}}, \sigma_P | P^{\text{OBS}}), \quad (2.10)$$

where  $P^{\text{DM}}$  is the output of the deterministic model and, again seeking a conservative ap-

proach, the variance  $\sigma_P$  prior distribution is set to Half-Normal distribution (the modulus of an ordinary normal distribution with mean zero) with a high value for scale parameter

$$\sigma_P \sim \mathcal{HN}(10^2). \quad (2.11)$$

The following sections detail each step of the UQ workflow.

### 2.7.1 Inverse Uncertainty Quantification

Once the Bayesian model is defined, the execution of the inference is configured through the adopted PPL, the PyMC.

```
with pm.Model() as model_pm:
    K_rv = pm.Uniform('K', lower=0.8*max(P_obs), upper=2*max(P_obs) )
    r_rv = pm.Uniform('r', lower=0.05, upper=1.5 )

    pop_rv = rand_pop(P_obs[0], K_rv, r_rv, t_obs)
    sigma = pm.HalfNormal('sigma', sd=error)

    P = pm.Normal('P', mu=pop_rv, sd=sigma, observed=P_obs)
```

With the NUTS sampler defined to draw  $10^4$  samples, the inference results is achieved and presented on Figure 2.10 with its sampled values to be diagnosed.

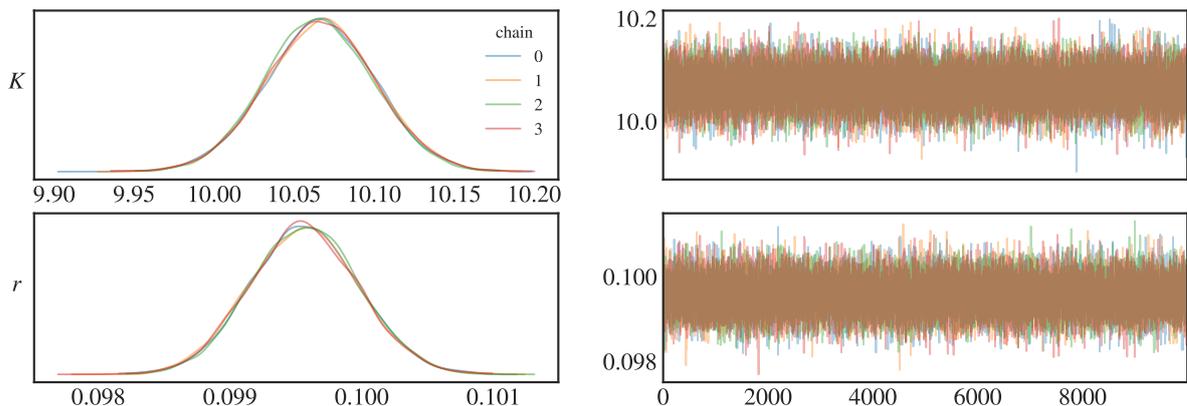


Figure 2.10: KDE of sampled PDFs and rank plot for the model input parameters.

*PyMC* as default iterates over 4 independent chains, the first column from Figure 2.10 presents these 4 chains via a kernel density estimates (KDE) almost at the same

shape. The chain KDE distribution together with the trace, at the right, indicates that the algorithm has converged with no difficulty.

In this example, the HDI of 99% is taken from the sample, as the approximate distributions have a high density around very specific values, i.e., with a low uncertainty. Once done, the results are ready to be carried to the uncertainty propagation.

### 2.7.2 Forward Uncertainty Quantification

Given the low computational cost to solve the problem (2.8), the classic MC method is employed to sample the approximate joint distributions of  $K$  and  $r$  and create a sample of the population growth along the time,  $P(t), t \in [0, 120]$ . First, the median of each sample is evaluated to get the best curve fit. Then 1024 observations are drawn via MC and evaluated with the simulator. Figure 2.11 presents some of the sampled curves.

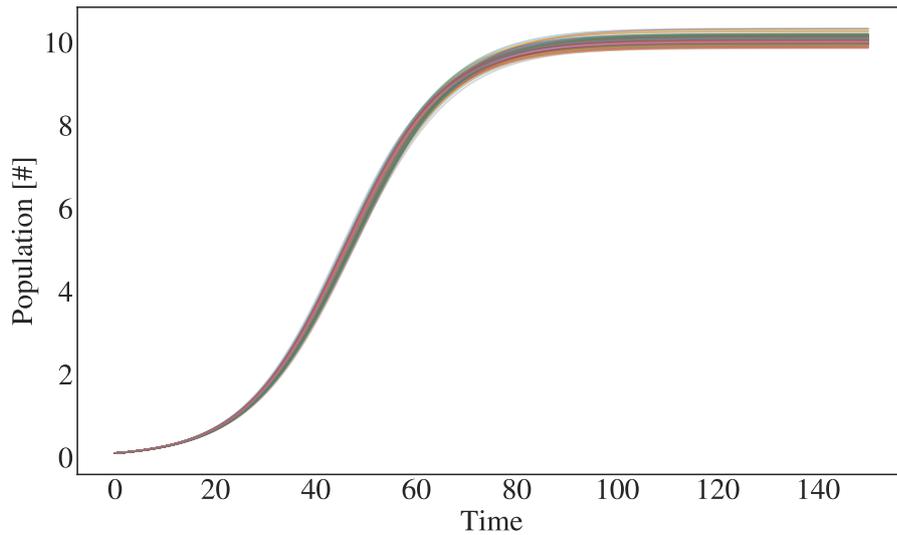


Figure 2.11: Propagated results via sampling from the posterior distributions approximations. *Notice that as the posterior PDFs of model parameters are estimated, the uncertainty propagation can be evaluated outside the data range.*

A prediction interval of confidence  $\gamma = 90\%$  is evaluated taking the 5<sup>th</sup> percentile and the 95<sup>th</sup> percentile. The best fit and the PI are presented in Figure 2.12. Notice that as the posterior PDFs of model parameters are estimated, the uncertainty propagation can be evaluated outside the data range.

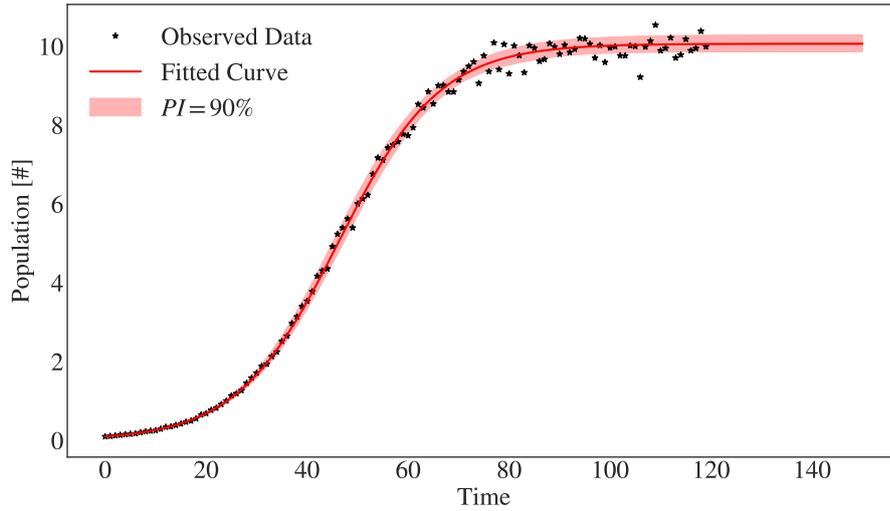


Figure 2.12: Propagation of uncertainties in the logistic growth model.

Notice that as the synthetic data was generated artificially with the noise respecting the adopted hypothesis of the error following a normal distribution, the PI is very restrained when compared to data. In other words, the proposed model accurately captures the phenomenon observed by the data (an obvious result, but one that implies in the correctness of the method).

### 2.7.3 Sensitivity Analysis

Finally, the main and total Sobol indices are evaluated over the observed time and presented in Figure 2.13. Some criticism of the model can be done now. The main and total indices for each parameter are almost exactly the same, which indicates that there is no interaction between the parameters affecting the sensitivity of the model. Also, it is possible to observe that the growth rate parameter,  $r$ , rules the sensitivity when population is growing, what makes sense as this parameter is designed to control this behavior. As greater the population size gets, lower is the growth rate influence and greater the influence of the carrying capacity,  $K$ . And when the population size reaches a stability, the carrying capacity becomes the responsible for all the variance of the model.

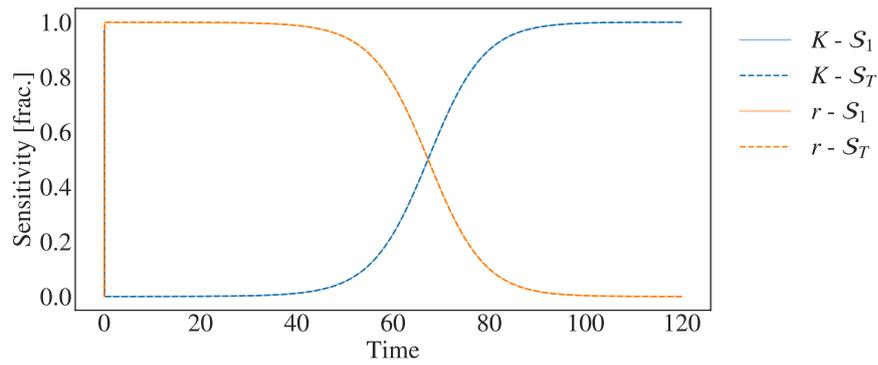


Figure 2.13: Main and total Sobol indices for  $K$  and  $r$  in respect to the population size along the time.

#### 2.7.4 Concluding Remarks

The inverse UQ retrieved both parameters,  $K$  and  $r$ , with a good accuracy and low uncertainty. The uncertainty propagated to the model respected the distribution of the generated synthetic data, as expected. Around  $t = 100$  it is possible to notice in Figure 2.12 that the model and its uncertainty reached a stable behavior even in a region where there is no data,  $t > 120$ . The behaviors observed in SA reaffirm the expected pattern of the model parameterization.

## 3 Foam Assisted Enhanced Oil Recovery

This chapter presents a case study (de Miranda et al., 2022) on the application of the presented UQ workflow for the parametric uncertainty estimation of a mathematical model used for EOR simulation using foam injection technique.

### 3.1 Background

In enhanced oil recovery (EOR) processes, foam injection reduces gas mobility and increases apparent viscosity, thus improving recovery efficiency. The reliability of simulators in the oil industry depends on the sensitivity of the outputs when subjected to input uncertainty. In this context, UQ is a key procedure while studying these physical systems.

A series of recent works done by Valdez et al. (VALDEZ et al., 2020; VALDEZ et al., 2021; VALDEZ et al., 2022) focused on UQ and SA using experimental steady-state data of foam quality in porous media. In particular, (VALDEZ et al., 2020) presented inverse and forward UQ analyses for different relative permeability models, together with a detailed SA. In the context of foam injection, (VALDEZ et al., 2021) presented the calibration of foam model parameters for a foam quality-scan experiment. Forward UQ and SA were carried out for the mathematical model used to characterize the effects of foam flow. Finally, in (VALDEZ et al., 2022) an improved procedure for foam parameter estimation was proposed when the non-Newtonian effects of foam flow are considered. In the context of Bayesian parameter inference, as is the case of these previous works, one has to provide prior knowledge not only about the parameters of interest but also for the variance of the experimental data. In these previous works, a conservative and direct choice for the prior of the variance was performed to estimate parametric uncertainty.

In this chapter will be presented a study to evaluate the effects of different prior distributions for the variance parameter of a Normal distribution assumed for experimental data. Assessing how the choice of the variance prior affects the estimated parameter distributions and how they affect the propagation of uncertainties to the outputs of the

model.

## 3.2 Mathematical Modeling

Consider the two-phase flow of water and gas in a porous medium subjected to the effects of foam to reduce the gas mobility and improve sweep efficiency. The general mathematical model for this problem is based on the principle of mass conservation for each phase,  $\alpha = \{w, g\}$ , which is given by

$$\frac{\partial}{\partial t} (\phi S_\alpha) + \frac{\partial}{\partial x} u_\alpha = 0, \quad \text{in } \Omega \times [0, T], \quad (3.1)$$

where  $S_\alpha$  and  $u_\alpha$  are the saturation and volumetric flux of phase  $\alpha$ . Writing  $u_\alpha$  as the Darcy velocity for the phase  $\alpha$ :

$$u_\alpha = -\lambda_\alpha \frac{\partial}{\partial x} p_\alpha, \quad \text{with} \quad \lambda_\alpha = \frac{K k_{r,\alpha}}{\mu_\alpha}, \quad \text{and} \quad \lambda_T = \sum_{\alpha} \lambda_\alpha, \quad (3.2)$$

where  $\lambda_\alpha$ ,  $k_{r,\alpha}$ ,  $\mu_\alpha$  and  $p_\alpha$  are respectively, the mobility, the relative permeability, the viscosity and the pressure of phase  $\alpha$ . Here,  $K$  is the absolute permeability of the porous medium and  $\lambda_T$  is the total mobility of the system.

### 3.2.1 Fractional Flow Theory

The fractional flow theory is a valuable tool for the understanding of the complex mechanisms involved in foam flow in porous media (DHOLKAWALA; SARMA; KAM, 2007; ROSSEN et al., 1999; ZHOU; ROSSEN, 1995). The fractional flow of a phase,  $f_\alpha$ , is the volumetric fraction of displacement in a direction composed by phase  $\alpha$  and can be defined as the ratio between the  $\lambda_\alpha$ -phase mobility and the total mobility, that is

$$f_\alpha = \frac{\lambda_\alpha}{\lambda_T} \implies \sum_{\alpha} f_\alpha = 1. \quad (3.3)$$

### 3.2.2 Relative Permeabilities

The Brooks-Corey model is used to describe the relative permeabilities of two-phase flow of water and gas (BROOKS; COREY, 1965). In the absence of foam, the relative permeabilities of water  $k_{r,w}$  and gas  $k_{r,g}$ , are given by:

$$k_{r,w} = k_{r,w}^0 \left( \frac{S_w - S_{wc}}{1 - S_{wc} - S_{gr}} \right)^{n_w}, \quad \text{and} \quad k_{r,g} = k_{r,g}^0 \left( \frac{S_g - S_{gr}}{1 - S_{wc} - S_{gr}} \right)^{n_g}, \quad (3.4)$$

where  $n_w$  and  $n_g$  are the Corey exponents for water and gas, respectively,  $k_{r,w}^0$  and  $k_{r,g}^0$  are the end-point relative permeabilities for water and gas, respectively,  $S_{wc}$  is the connate water saturation, and  $S_{gr}$  is the residual gas saturation.

In the absence of foam, it is possible to obtain these values from experimental data in terms of apparent viscosity and fractional flow

$$k_{r,w} = \frac{(1 - f_g)\mu_w}{\mu_{app}}, \quad \text{and} \quad k_{r,g} = \frac{f_g\mu_g}{\mu_{app}}. \quad (3.5)$$

For a two-phase flow in the presence of foam, the relation for water permeability is still the same, but the gas mobility is affected by its presence (EFTEKHARI; FARAJZADEH, 2017). In this work, an implicit-texture model given by the CMG-STARS model is used to represent the effects of foam on the two-phase flow.

### 3.2.3 Foam Modeling

The CMG-STARS model describes the foam effects in two-phase flow using the mobility reduction factor ( $MRF$ ) term. This term can describe the effects of surfactant concentration, water and oil saturations, shear-thinning, and other effects on the flow. This term is introduced in the apparent viscosity, or the inverse of total relative mobility, as follows

$$\mu_{app} = \lambda_T^{-1} = \left( \lambda_w + \frac{\lambda_g}{MRF} \right)^{-1}, \quad (3.6)$$

to take into account that the mobility of the gas phase is affected by the foam via the  $MRF$  term. The fractional gas flow is also redefined as follows:

$$f_g = \frac{\lambda_g}{MRF \left( \lambda_w + \frac{\lambda_g}{MRF} \right)} = \frac{\lambda_g}{MRF} \mu_{app}. \quad (3.7)$$

Finally, the mobility reduction factor  $MRF$  is defined as

$$MRF = 1 + fmmob \prod_i F_i. \quad (3.8)$$

A wide range of effects can be included in the  $MRF$  through the  $F_i$  terms of the CMG-STARS model. This work only considers the  $F_2$  term that describes the effects of water saturation on foam given by

$$F_2 = \frac{1}{2} + \frac{1}{\pi} \arctan(sfbet(S_w - SF)), \quad (3.9)$$

where  $\Theta = \{fmmob, sfbet, SF\}$  are STARS model parameters, which describe, respectively, the maximum gas mobility reduction, the sharpness of the transition between high-quality and low-quality regime, and the critical water saturation around which foam collapses.

### 3.2.4 Experimental Data

To carry out the studies of this work, the experimental dataset of foam quality-scan reported in (VALDEZ et al., 2021) will be investigated using the Corey parameters fitted by (MOHAMED; NASR-EL-DIN et al., 2012) to a high permeability Indiana Limestone. The parameters obtained are given in Table 3.1.

Parameters	$S_{wc}$	$S_{gr}$	$n_w$	$n_w$	$k_{r,w}^0$	$k_{r,g}^0$
Values	0.4	0.293	2.98	0.96	0.302	0.04

Table 3.1: Relative permeability parameters for datasets.

(TAO; WATSON, 1984) studied errors in relative permeabilities estimated from displacement experiments under different operating conditions. The results showed that

the error ranged from 0.2% to 15%. In most cases, the error is less than 5%, except for higher errors near residual water saturation. The fitting of the permeability model parameters in a reliable way and uncertainty estimation is also a challenging problem that will not be tackled in this work since the main focus is on the CMG-STARs model parameters.

Figure 3.1 presents the experimental data for pressure drop at different foam quality rates ( $f_g$ ) as a function of time. At the end of each foam quality stage, it is assumed the pressure drop reached the steady-state, and this is verified in almost all foam quality ranges. However, for  $f_g = 0.5$  it is difficult to identify this behavior. This observation is reflected in the stationary data presented in Figure 3.2 in terms of apparent viscosity obtained from experimental data using the following relation  $\mu_{app} = \frac{K}{v} \frac{\Delta p}{L}$ , where  $v$  is the superficial velocity and  $L$  the core sample length. Figure 3.2 also presents an example of the CMG-STARs model, defined by equations (3.4) - (3.9), fitted (solid line) to experimental data.

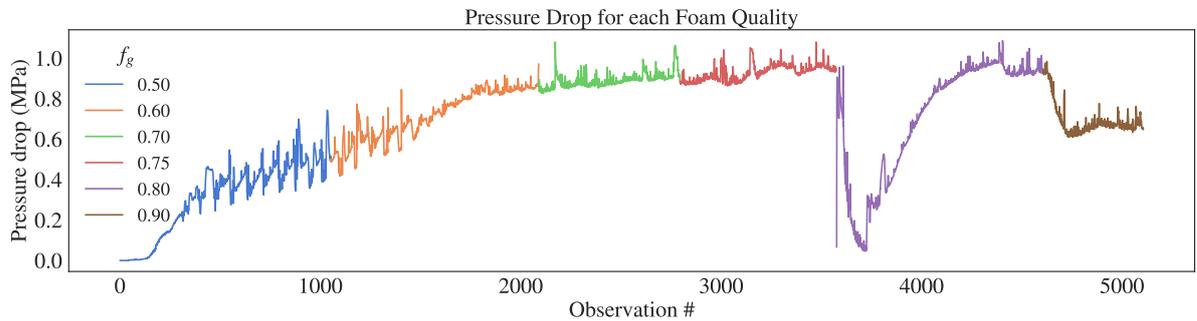


Figure 3.1: Unsteady-State (USS) experimental dataset for pressure drop in a Indiana Limestone for each foam quality.

### 3.2.5 Hierarchical Modeling and Noninformative Priors

In Bayesian statistics, some practitioners support if there is no prior knowledge of the ways an event can occur, the event will occur equally likely in any way. This is known as the Principle of Insufficient Reason (PIR) enunciated by Jakob Bernoulli (JAYNES, 2003). (BOX; TIAO, 1992) define a noninformative prior as a prior that provides little information about the studied variable. (GELMAN et al., 1995) advocates that there is no truly noninformative prior, in the sense that any prior includes some information. But

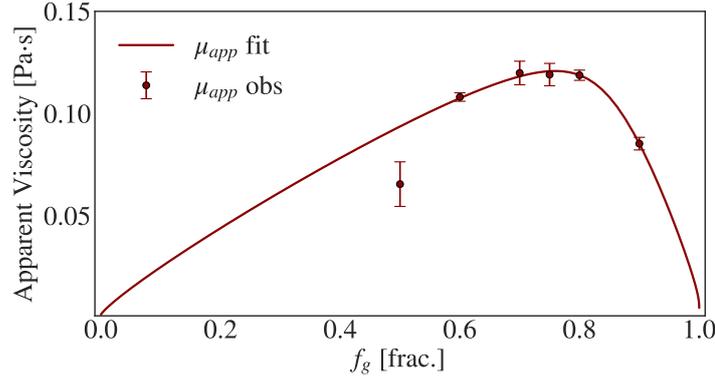


Figure 3.2: Steady-State (SS) experimental dataset for foam quality scans, fitted via an optimization algorithm to illustrate the composed by equations (3.4) - (3.9).

adopting the name coined by the pioneers, in mathematical terms, following the PIR and defining a noninformative prior can be done through the uniform distribution over the feasible interval.

For the CMG-STARS foam parameters, the choice is based on previous works from (VALDEZ et al., 2022; VALDEZ et al., 2021), that is:

$$fmmob \sim \mathcal{U}[0.0, 1000], \quad SF \sim \mathcal{U}[S_{wc}, 1 - S_{gr}], \quad \text{and} \quad sfbet \sim \mathcal{U}[10, 1000]. \quad (3.10)$$

Having the parameters priors defined in Eq. (3.10), the model is completed assuming the data follows a Normal distribution around the output of the deterministic model of apparent viscosity, with an associated error/variance. In statistical notation, this is written as

$$\mu_{app} \sim \mathcal{N}(\mu_{app}^{DM}, \sigma_{\mu_{app}} | \mu_{app}^{obs}), \quad (3.11)$$

with

$$\mu_{app}^{DM} = \mu_{app}^{DM}(fmmob, SF, sfbet | f_g^{obs}), \quad (3.12)$$

as defined by equations (3.4) to (3.9) that composes our deterministic model (DM) to be randomized by the Probability Density Functions (PDF) of  $fmmob$ ,  $SF$  and  $sfbet$ . Here  $\sigma_{\mu_{app}}$  is the error committed in the approximation, which comprehends aleatoric and epistemic uncertainties, and can also be modeled as a random variable.

In the following subsections, different choices for the standard deviation of the apparent viscosity are presented in detail, considering the data of standard deviations for the apparent viscosity  $\sigma_{\mu_{app}}^{obs}$  (as represented by the bars in Figure 3.2).

### 3.2.6 On the choice of prior distributions for variance parameter

Three different experiments were carried out to evaluate how a more informative prior for the scale parameter of the data distribution contributes to the inference process. A more conservative family of distributions is employed to ensure that bias in prior knowledge does not compromise the extraction of information from observed data. Finally, a proposal for weighting the deviations using the property of the scale family of probability distributions is used to bring more flexibility to the Bayesian model.

**Experiment I (Maximum Variance)** A common choice of prior for variance parameters expecting a conservative approach is the default half-normal with its scale parameter set to a high value. So, the first approach studied employs a prior parameterized by the maximum observed value, which in probabilistic notation is defined as

$$\sigma_{\mu_{app}} \sim \mathcal{HN} \left( \max \left( \sigma_{\mu_{app}}^{obs} \right) \right). \quad (3.13)$$

In particular, note that the maximum error observed, denoted by  $\max(\sigma_{\mu_{app}}^{obs})$ , occurs at  $f_g = 0.5$ , as shown in Figure 3.2. Its value is significantly higher than the other errors, as can be seen, which guarantees that the sample deviation will not be underestimated.

**Experiment II (Maximum Likelihood Estimation (MLE) of Variance)** In the second Experiment, a stronger prior choice is sought to test the influence of top-level parameter  $\sigma$  in the inference of the parameters of interest. (GELMAN, 2006) suggests that if a more informative prior for variance is desired, the half-Student-t (or just half-t) distribution might be recommended.

Following (GELMAN, 2006), the present experiment uses a maximum likelihood estimation (MLE) for the scale parameter  $\hat{\sigma}_{app}$  over the standard deviation of the experimental data, i.e., the parameter value that maximizes the probability of the observed

data. Thus, the MLE is set as the half-t scale parameter. The half-t prior also requires another parameter, the degrees of freedom parameter  $\nu$ , which concerns the number of observations  $n$ , by the relation  $\nu = n - 1$ . The degrees of freedom are the number of available observations  $n$  minus one, as shown in Figure 3.2 by the number of data points. To summarize, in this experiment the following prior is used:

$$\sigma_{\mu_{app}} \sim \mathcal{HT}(n - 1, \hat{\sigma}_{app}). \quad (3.14)$$

Figure 3.3 compares a half-normal and a half-t using the same value for the scale parameter. Notice that, due to the fatter tail, the half-t has more probability of extreme values than the half-normal distribution.

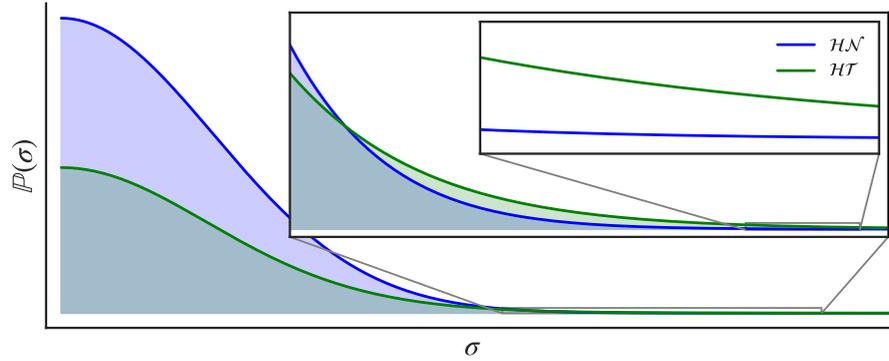


Figure 3.3: Comparison of a half-normal and a half-t distribution with a same scale parameter, and degrees of freedom,  $\nu = n - 1$ . The first zoom shows the point where the half-t PDF becomes superior than the half-normal PDF. The second zoom shows the fatter tail preserved.

**Experiment III (Weighting Variance)** In parameter estimation, when the information about the error made in the measurement is available, that is, the deviation from the observation, it is possible to weight each point according to its variance: the greater the observed variance, the less weight it will receive in the fitting process. In the context of parameter estimation for foam flow models, weighting the data points has been applied with least-squares methods (MA et al., 2014). However, in Bayesian inference packages, such as in the PyMC3, such weighting strategy is not trivial as when using non-linear least-squares methods.

In the present experiment, a weighting strategy is proposed that relies on the no-

tion of distributions within the scale family, such as half-Normal and half-t. The strategy proposed uses the structure provided by PyMC3 and rewrites the distribution of the data, equation (3.11), using a random vector of size  $n$  for the random variable  $\sigma_{\mu_{app}}$  according to its observed deviations at each  $f_g$ . This choice can be described as

$$\mu_{app} \sim \mathcal{N}\left(\mu_{app}^{DM}, \sigma_{\mu_{app}}^{obs} \cdot \sigma_{\mu_{app}}\right). \quad (3.15)$$

Now the prior adopted for  $\sigma_{\mu_{app}}$  is a standard PDF, since it will be scaled for each observed deviation (as reported in Figure 3.2). The family of distribution will be the same from Experiment II, expecting to be more conservative.

The degrees of freedom parameter is selected considering how the sampling standard deviations at each  $f_g$  point are calculated. Figure 3.1 shows the pressure drop as a function of time for different observed ranges of  $f_g$ . The calculation for the steady-state dataset (Figure 3.2) is done by taking the last  $m$  experimental points for each foam quality  $f_g$  level, where  $m$  is chosen properly to capture the reached steady-state behavior. Therefore, the following choice is considered:

$$\sigma_{\mu_{app}} \sim \mathcal{HT}(m - 1). \quad (3.16)$$

Figure 3.4 presents a graphical representation of the random vector of PDFs with its components associated with each point in Figure 3.2. It is interesting to observe that the PDF associated with the data at  $f_g = 0.5$  shows more probability for extreme values due to its larger standard deviation (see Figure 3.2), as defined by equation (3.15).

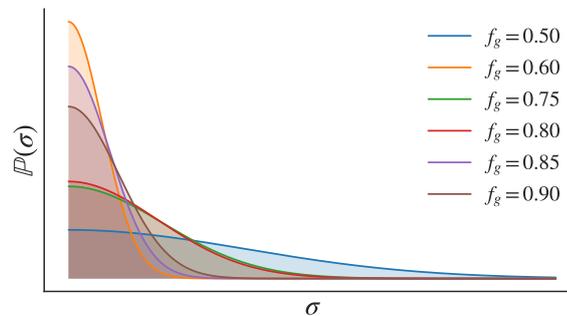


Figure 3.4: Prior distributions set for the observed errors associated with each  $f_g$  point presented in Figure 3.2. Note that larger observed deviation correspond to flatter associated prior distribution.

**Computational Experiments Setting** With the experimental data (Figure 3.2) and the model composed by equations (3.10), (3.11), and one choice for the prior distribution of variance given by (3.13) or (3.14) or (3.16), again the PyMC with the NUTS sampler is employed for Bayesian inference process.  $5 \times 10^4$  samples were drawn for each randomized parameter, and  $10^4$  samples were discarded from each of the final chains. The joint of these four chains describes a sample of the posterior distribution for each parameter.

## 3.3 Results

### 3.3.1 Experiment I - Maximum Variance

Figure 3.5 illustrates the posterior distribution of each parameter obtained with the choice of prior distribution for variance as described for the first experiment. The KDE for the PDFs of foam parameters looks similar to the ones achieved by (VALDEZ et al., 2021). The parameter *sfbet* shows more asymmetry and less kurtosis than the other ones exactly as happened in Valdez experiment with only STARS foam parameters.

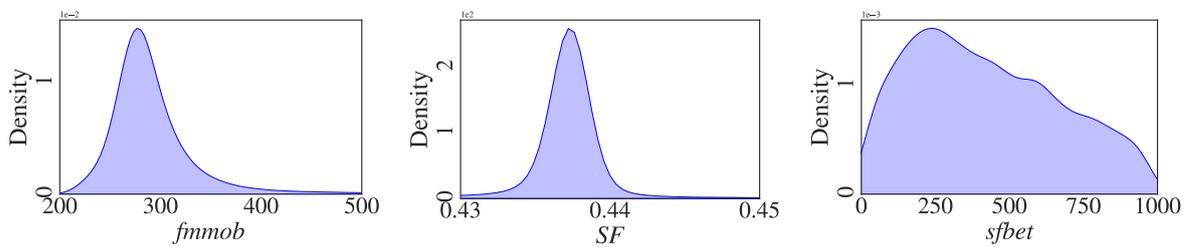


Figure 3.5: Kernel Density Estimate (KDE) plot for the posterior distributions for each parameter at Experiment I.

Once the posterior distributions are reached, the study proceeds to propagate the uncertainty of the samples obtained using the MC method. Figure 3.6 shows the average value sampled for the curves and a prediction interval with 90% uncertainty (represented by the shaded region). The uncertainty propagated from the posteriors is similar to the results presented by (VALDEZ et al., 2021), where uncertainty in the high-quality regime is more evident, and the point for  $f_g = 0.5$  is out of the prediction interval.

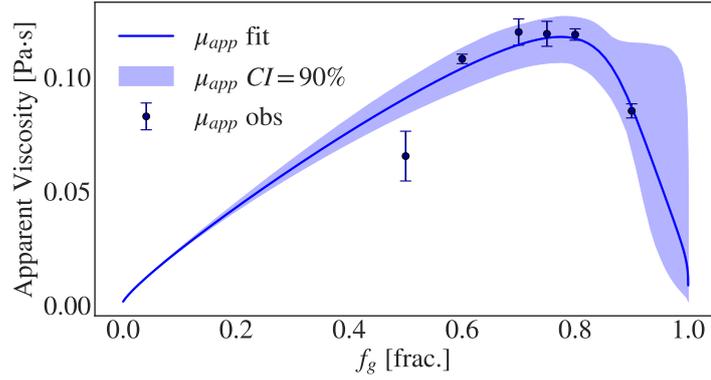


Figure 3.6: Uncertainty propagation with 90% prediction interval for Experiment 1.

### 3.3.2 Experiment II - Maximum Likelihood Estimation (MLE) of Variance

With the model given by (3.10), (3.11) and (3.14), the Bayesian inference in the same setup as the Experiment 1 is executed. Figure 3.7 shows the KDE for each posterior distribution achieved.

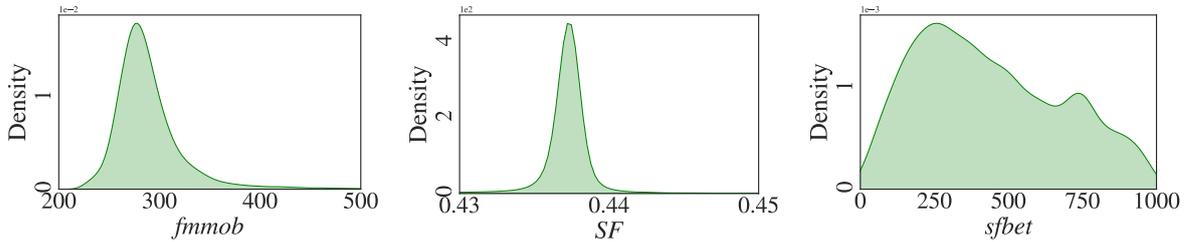


Figure 3.7: Kernel Density Estimate (KDE) plot for the posterior distributions for each parameter at Experiment II.

Figure 3.7 shows the propagated uncertainties to the apparent viscosity. The results of this experiment present some reduction of the uncertainties in the high-quality regime.

### 3.3.3 Experiment III - Weighting the Variance

The posterior distributions obtained after the Bayesian inference in the third experiment, where the model is described by (3.10), (3.11) and (3.16), are reported in Figure 3.9. One can clearly observe that the posterior distributions obtained in this case provide estimates with higher fidelity and lower uncertainties. This fact is more evident for the

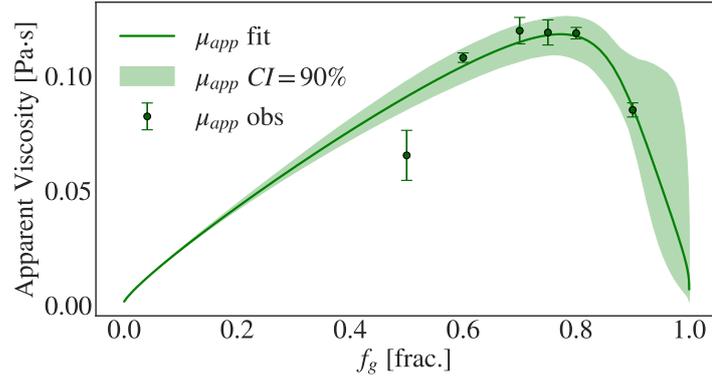


Figure 3.8: Uncertainty propagation with 90% prediction interval for Experiment II.

$sfbet$  parameter whose PDF indicates a median value around 300 and which tails out to zero, in contrast with the previous two cases.

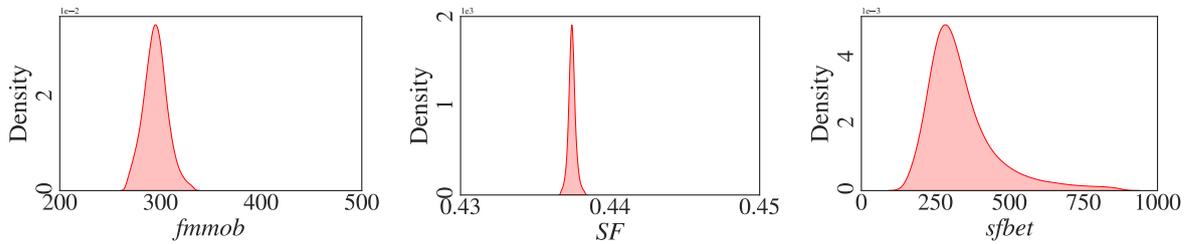


Figure 3.9: Kernel Density Estimate (KDE) plot for the posterior distributions for each parameter at Experiment III.

The uncertainty propagation in this experiment is far more controlled as shown in Figure 3.10. This is due to the better defined and less wide posterior distribution achieved for  $sfbet$  parameter, which controls the transition between low- and high-quality regimes.

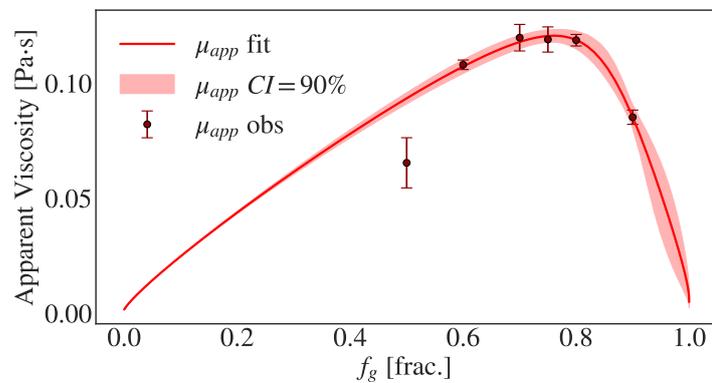


Figure 3.10: Uncertainty propagation with 90% prediction interval for Experiment III.

### 3.3.4 Experiments Comparison

Table 3.2 presents the median and standard deviation from the PDFs of each parameter in order to compare the results among the three experiments. The results show a good agreement in the median values of the parameters, and also show through the standard deviation that the Experiment III delivers less uncertainty in the parametric range.

	Median $\pm$ Standard Deviation		
	$fmmob$ ( $\times 10^2$ )	$SF$ ( $\times 10^{-1}$ )	$sfbet$ ( $\times 10^2$ )
Experiment I	$3.041 \pm 0.733$	$4.359 \pm 0.107$	$4.615 \pm 2.43$
Experiment II	$2.876 \pm 0.470$	$4.391 \pm 0.0892$	$4.698 \pm 2.56$
Experiment III	$3.049 \pm 0.222$	$4.367 \pm 0.0753$	$3.038 \pm 0.894$

Table 3.2: Median and standard deviation of the estimated PDFs for each parameter in each experiment conducted.

Figure 3.11 presents a comparison of the PDFs of each parameter obtained in the three experiments. One can clearly observe the fact that Experiment III delivers better defined PDFs with less uncertainty, especially for the  $sfbet$  parameter.

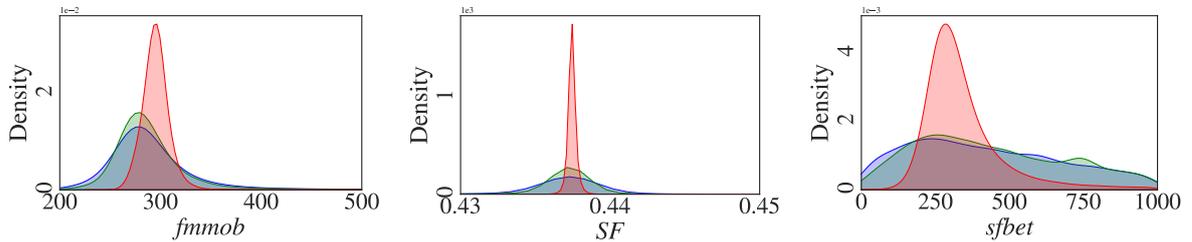


Figure 3.11: Comparison of KDEs for each parameter obtained by the different experiments conducted.

A further comparison of the experiments' results considers the error between experimental data and the data simulated with median estimated parameters presented in Table 3.2. The error is measured using the mean absolute percentage error (MAPE) and median absolute percentage error (MDAPE). It is shown in Table 3.3, that unweighted estimators (Exp. I and II), present almost the same performance as the weighted ones (Exp. III) when MAPE is used. However, with the MDAPE it is possible to see the improvements of the proposed strategy.

	MAPE	MDAPE
Experiment I	$3.94 \times 10^{-2}$	$1.34 \times 10^{-2}$
Experiment II	$3.87 \times 10^{-2}$	$1.21 \times 10^{-2}$
Experiment III	$3.85 \times 10^{-2}$	$6.71 \times 10^{-3}$

Table 3.3: Comparison of the results among the three experiments using MAPE and MDAPE.

### 3.3.5 Sensitivity Analysis

The main and total Sobol indices with respect to apparent viscosity ( $\mu_{app}$ ) are observed using a 90% HDI of the posterior distribution obtained in Experiment III, to understand the influence of each parameter. The results presented in Figures 3.12 and 3.13 show that the variance of apparent viscosity strongly influences  $fmmob$  in the low-quality regime. Around the low- and high-quality transition region, the  $sfbet$  parameter presents a significant impact, followed by a brief interval where  $SF$  impacts the apparent viscosity. This result agrees with those presented by (VALDEZ et al., 2021). However, it captures in a more precise way the influences near the transition region, where  $SF$  has a brief interval of significant contribution followed by the dominance of  $sfbet$  in the high-quality regime.

It is also possible to modify the QoI, keeping the apparent viscosity output but now in terms of the water saturation,  $\mu_{app}(S_w)$ . Notice that the transition region respects exactly the point  $S_w = SF$ . Going further, when comparing the figures, there are indications that lower water saturations typically appear in the high-quality region (dominated by  $sfbet$ ). After the foam abrupts,  $S_w > SF$ ,  $fmmob$  rules the sensitivity of the apparent viscosity.

## 3.4 Experiment Conclusion

The usage of standard deviation information at each foam quality point as a weighting factor in Bayesian inference improved foam model reliability. It was shown that it significantly improved the estimation of the  $sfbet$  parameter, and as a consequence, it substantially reduced the uncertainty propagated to the high-quality foam regime. In addition, the sensitivity analysis results are consistent with the physics of the model for

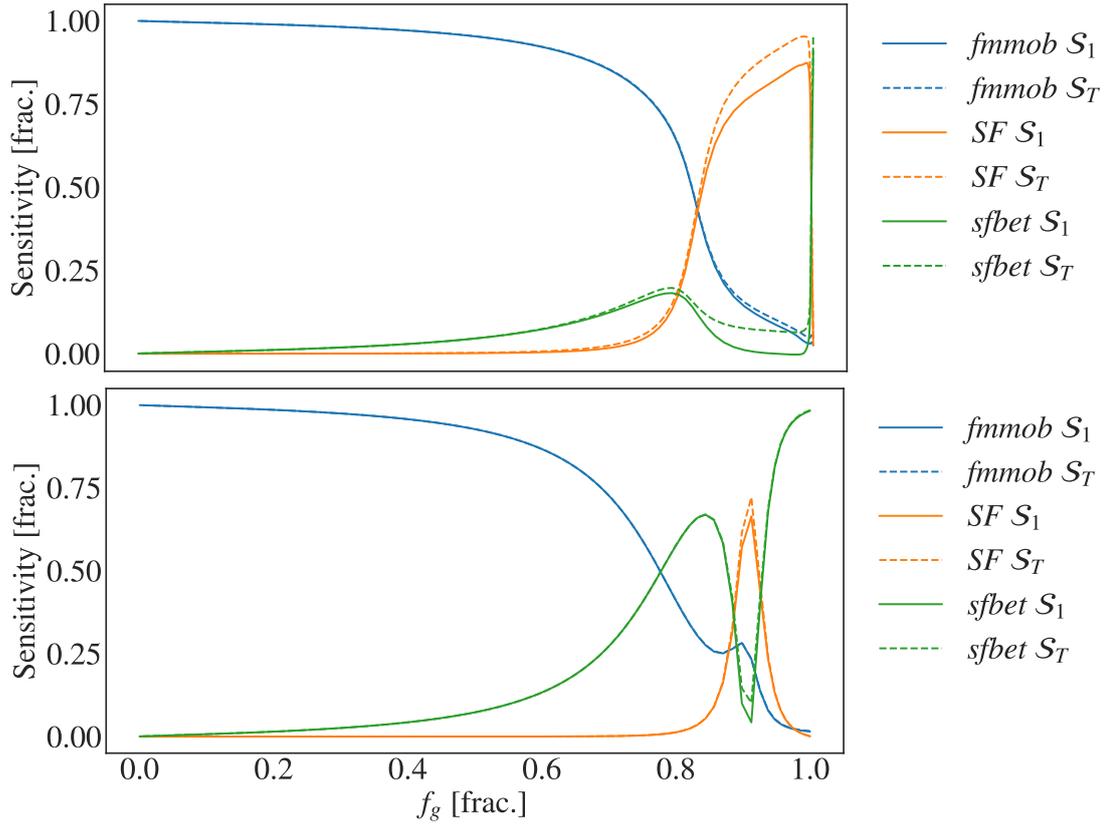


Figure 3.12: Sobol Indices for the apparent viscosity employing the 90% HDI of PDFs from Experiments I and III. As presented in Eq. (2.7),  $\mathcal{S}_1$  denotes the first-order Sobol index, and  $\mathcal{S}_T$  the total Sobol index.

the underlying phenomenon. The *fmmob* parameter remains one of the key parameters impacting the apparent viscosity. The improved estimation of the *sfbet* parameter with a better defined posterior distribution contributed to highlighting its role in the sensitivity analysis carried out.

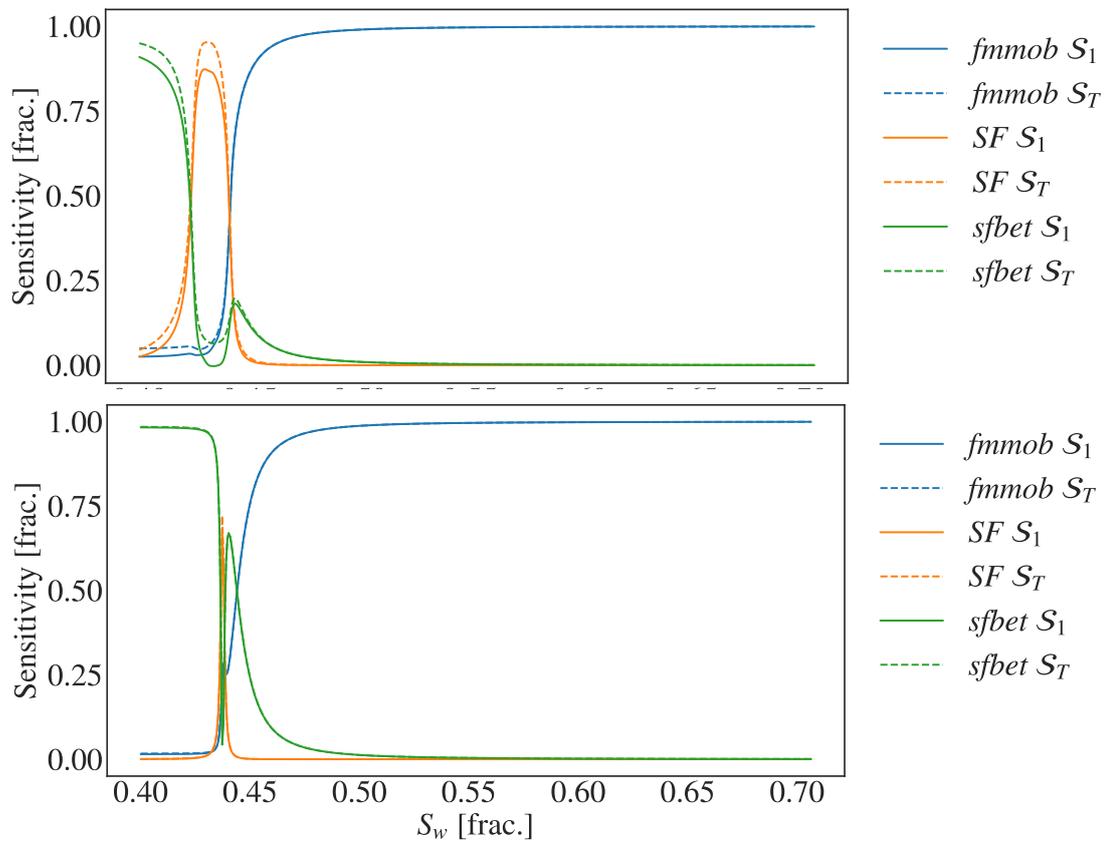


Figure 3.13: Sobol Indices for the apparent viscosity employing the 90% HDI of PDFs from Experiment I and III.

## 4 Conclusions

This work presented a general approach to quantify uncertainties in simulators seeking to detail each step of the workflow. Techniques commonly used in each workflow step have been presented and remarked. A simple illustrative case was briefly investigated to prepare the reader with the concepts and possibilities of the techniques employed in this work.

Inverse and Forward UQ analyses were performed for a foam-assisted EOR model considering the parameters of the CMG-STARs foam model with different choices for variance parameter of the hierarchical model's top-level distribution, i.e., the hypothesis for data distribution. The proposed approach worked with statistics over the unsteady-state dataset and improved the parameter estimation obtaining estimates with reduced uncertainty, sensitivity analysis and model reliability.

### 4.1 Limitations and Future Works

As the number of parameters grows, the computational cost in the inverse UQ step grows dramatically. Inverse UQ usually deals with ill-posed problems; thus, delicate issues, such as strong non-uniqueness, arise in estimating parameters and uncertainties. A more informative prior distribution in these cases is desired, but this can't be an arbitrary choice, as presented in this work. Transfer learning strategies can bring information from other computational studies when dealing with more complex models drive a the inference to a better result (PAN; YANG, 2009).

In the context of EOR, when considering more factors that influence the foam, such as non-Newtonian effects or even the presence of oil (a natural next step in the problem), the difficulty of finding the median of a set of parameters that describe the observed data.

As mentioned in the previous section, a natural next step is considering the oil phase in the mathematical model. To that end, three-phase flow data must be used within

the pipeline to include another term in the CMG-STARS model function describing the effects of oil saturation. In particular, this term is given by

$$F_3 = \left( \frac{(fmoil - S_o)}{(fmoil - floil)} \right)^{epoil},$$

where  $fmoil, floil \in [0, 1]$  are the critical oil saturation and the lower oil saturation values, respectively.

In addition, one may desire computational expensive QoIs. For this purpose, emulators should be used to reduce the computational cost in the propagation of uncertainties and sensitivity analysis. It is also an author's interest to explore new techniques of sensitivity analysis, such as the one already mentioned in this text, the Shapley effects.

## Bibliography

- ALVIN, K. et al. Uncertainty quantification in computational structural dynamics: a new paradigm for model validation. In: CITESEER. *Society for Experimental Mechanics, Inc., 16 th International Modal Analysis Conference*. [S.l.], 1998. v. 2, p. 1191–1198.
- BALLESTEROS, G. C. et al. Bayesian hierarchical models for uncertainty quantification in structural dynamics. *Vulnerability, uncertainty, and risk: Quantification, mitigation, and management*, American Society of Civil Engineers (ASCE) Reston, VA, v. 162, p. 1615–1624, 2014.
- BERG, S.; UNSAL, E.; DIJK, H. Non-uniqueness and uncertainty quantification of relative permeability measurements by inverse modelling. *Computers and Geotechnics*, Elsevier, v. 132, p. 103964, 2021.
- BOX, G. E.; TIAO, G. C. *Bayesian inference in statistical analysis*. [S.l.]: John Wiley & Sons, 1992. v. 40.
- BROOKS, R. H.; COREY, A. T. *Hydraulic properties of porous media*. [S.l.]: Colorado State University, 1965.
- BROOKS, S. et al. *Handbook of markov chain monte carlo*. [S.l.]: CRC press, 2011.
- CASELLA, G.; BERGER, R. L. *Statistical Inference*. [S.l.]: Cengage Learning, 2021. ISBN 978-0-357-75313-2.
- CEDRO, J. B. *Influência da pressão capilar em escoamentos bifásicos com espuma em meios porosos*. Federal University of Juiz de Fora, 2020. Disponível em: <https://repositorio.ufjf.br/jspui/handle/ufjf/12422>.
- (CMG)., C. M. G. *STARS Users Manual; Version 2019.10*. [S.l.]: CMG Ltd Calgary, Canada, 2019.
- COUNCIL, N. R. et al. *Assessing the reliability of complex models: mathematical and statistical foundations of verification, validation, and uncertainty quantification*. [S.l.]: National Academies Press, 2012.
- de Miranda, G. B. et al. An improved approach for uncertainty quantification based on steady-state experimental data in foam-assisted enhanced oil recovery. In: EAGE - EUROPEAN ASSOCIATION OF GEOSCIENTISTS AND ENGINEERS. *ECMOR 2022: The 18th European Conference on the Mathematics of Oil Recovery*. [S.l.], 2022. Accepted/To Appear.
- DHOLKAWALA, Z. F.; SARMA, H.; KAM, S. Application of fractional flow theory to foams in porous media. *Journal of Petroleum Science and Engineering*, Elsevier, v. 57, n. 1-2, p. 152–165, 2007.
- DUANE, S. et al. Hybrid monte carlo. *Physics letters B*, Elsevier, v. 195, n. 2, p. 216–222, 1987.

- EFTEKHARI, A. A.; FARAJZADEH, R. Effect of foam on liquid phase mobility in porous media. *Scientific reports*, Nature Publishing Group, v. 7, n. 1, p. 1–8, 2017.
- GELMAN, A. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*, International Society for Bayesian Analysis, v. 1, n. 3, p. 515–534, 2006.
- GELMAN, A. et al. *Bayesian data analysis*. [S.l.]: Chapman and Hall/CRC, 1995.
- GELMAN, A. et al. Bayesian data analysis third edition (updated 15 february 2021). p. 677, 2021.
- GHANEM, R. G.; SPANOS, P. D. Stochastic finite element method: Response statistics. In: *Stochastic finite elements: a spectral approach*. [S.l.]: Springer, 1991. p. 101–119.
- GRAMACY, R. B. *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*. [S.l.]: Chapman and Hall/CRC, 2020.
- HARRIS, C. R. et al. Array programming with NumPy. *Nature*, Springer Science and Business Media LLC, v. 585, n. 7825, p. 357–362, set. 2020. Disponível em: <https://doi.org/10.1038/s41586-020-2649-2>.
- HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. v. 57, n. 1, p. 97–109, 1970. ISSN 0006-3444. Disponível em: <https://doi.org/10.1093/biomet/57.1.97>.
- HOFFMAN, M. D.; GELMAN, A. et al. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, v. 15, n. 1, p. 1593–1623, 2014.
- JAYNES, E. T. *Probability theory: The logic of science*. [S.l.]: Cambridge university press, 2003.
- KRUSCHKE, J. *Doing Bayesian Data Analysis: A Tutorial with R, Jags, and Stan*. 2nd revised ed.. ed. [S.l.]: Academic Press, 2015. ISBN 978-0-12-405888-0.
- KUMAR, R. et al. Arviz a unified library for exploratory analysis of bayesian models in python. *Journal of Open Source Software*, The Open Journal, v. 4, n. 33, p. 1143, 2019. Disponível em: <https://doi.org/10.21105/joss.01143>.
- LEE, J.; SUNG, W.; CHOI, J.-H. Metamodel for efficient estimation of capacity-fade uncertainty in li-ion batteries for electric vehicles. *Energies*, MDPI, v. 8, n. 6, p. 5538–5554, 2015.
- LI, J. et al. *Uncertainty quantification of two-phase flow in porous media via coupled-TgNN surrogate model*. arXiv, 2022. Disponível em: <https://arxiv.org/abs/2205.14301>.
- LOHR, S. L. *Sampling: design and analysis*. [S.l.]: Chapman and Hall/CRC, 2021.
- MA, K. et al. Non-uniqueness, numerical artifacts, and parameter sensitivity in simulating steady-state and transient foam flow through porous media. *Transport in porous media*, Springer, v. 102, n. 3, p. 325–348, 2014.
- MARTIN, O.; WIECKI, T. *Bayesian analysis with Python: introduction to statistical modeling and probabilistic programming using PyMC3 and ArviZ*. Second edition. [S.l.]: Packt Publishing Limited, 2018. ISBN 978-1-78934-165-2.

- MARTIN, O. A.; KUMAR, R.; LAO, J. *Bayesian Modeling and Computation in Python*. Boca Raton: [s.n.], 2021. ISBN 978-0-367-89436-8.
- MCELREATH, R. *Statistical rethinking: A Bayesian course with examples in R and Stan*. [S.l.]: Chapman and Hall/CRC, 2020.
- METROPOLIS, N. et al. Equation of state calculations by fast computing machines. v. 21, n. 6, p. 1087–1092, 1953. ISSN 0021-9606. Publisher: American Institute of Physics. Disponível em: <https://aip.scitation.org/doi/10.1063/1.1699114>.
- METROPOLIS, N.; ULAM, S. The monte carlo method. *Journal of the American statistical association*, Taylor & Francis, v. 44, n. 247, p. 335–341, 1949.
- MOHAMED, I.; NASR-EL-DIN, H. et al. Formation damage due to co2 sequestration in deep saline carbonate aquifers. In: SOCIETY OF PETROLEUM ENGINEERS. *SPE International Symposium and Exhibition on Formation Damage Control*. [S.l.], 2012.
- PAN, S. J.; YANG, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, IEEE, v. 22, n. 10, p. 1345–1359, 2009.
- RIBEIRO, L. et al. A workflow for uncertainty quantification of numerical models for foam-based eor: emulators, bayesian inference, forward uncertainty quantification, and sensitivity analysis. In: IBP - INSTITUTO BRASILEIRO DE PETRÓLEO E GÁS. *Rio Oil & Gas Expo And Conference, 2022. — ISSN 2525-7579*. [S.l.], 2022. Accepted/To Appear.
- ROSSEN, W. et al. Simplified mechanistic simulation of foam processes in porous media. *SPE Journal*, OnePetro, v. 4, n. 03, p. 279–287, 1999.
- RUDE, U. et al. Research and education in computational science and engineering. *Siam Review*, SIAM, v. 60, n. 3, p. 707–754, 2018.
- SALTELLI, A. et al. *Global sensitivity analysis: the primer*. [S.l.]: John Wiley & Sons, 2008.
- SALVATIER, J.; WIECKI, T. V.; FONNESBECK, C. Probabilistic programming in python using PyMC3. *PeerJ Computer Science*, PeerJ, v. 2, p. e55, apr 2016. Disponível em: <https://doi.org/10.7717/peerj-cs.55>.
- SEDEHI, O.; PAPADIMITRIOU, C.; KATAFYGIOTIS, L. S. Data-driven uncertainty quantification and propagation in structural dynamics through a hierarchical bayesian framework. *Probabilistic Engineering Mechanics*, Elsevier, v. 60, p. 103047, 2020.
- SOBOL', I. M. On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe modelirovanie*, Russian Academy of Sciences, Branch of Mathematical Sciences, v. 2, n. 1, p. 112–118, 1990.
- SOBOL, I. M. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation*, Elsevier, v. 55, n. 1-3, p. 271–280, 2001.
- TAO, T.; WATSON, A. Accuracy of jbn estimates of relative permeability: Part 1—error analysis. *Society of Petroleum Engineers Journal*, OnePetro, v. 24, n. 02, p. 209–214, 1984.

TEAM, T. pandas development. *pandas-dev/pandas: Pandas*. Zenodo, 2020. Disponible en: <https://doi.org/10.5281/zenodo.3509134>.

VALDEZ, A. R. et al. Uncertainty quantification and sensitivity analysis for relative permeability models of two-phase flow in porous media. *Journal of Petroleum Science and Engineering*, Elsevier, v. 192, p. 107297, 2020.

VALDEZ, A. R. et al. Assessing uncertainties and identifiability of foam displacement models employing different objective functions for parameter estimation. *Journal of Petroleum Science and Engineering*, Elsevier, p. 110551, 2022.

VALDEZ, A. R. et al. Foam-assisted water–gas flow parameters: From core-flood experiment to uncertainty quantification and sensitivity analysis. *Transport in Porous Media*, Springer, p. 1–21, 2021.

WIENER, N. The homogeneous chaos. *American Journal of Mathematics*, JSTOR, v. 60, n. 4, p. 897–936, 1938.

XIU, D.; KARNIADAKIS, G. E. The wiener–askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, SIAM, v. 24, n. 2, p. 619–644, 2002.

ZHOU, Z.; ROSSEN, W. Applying fractional-flow theory to foam processes at the limiting capillary pressure. *SPE Advanced Technology Series*, OnePetro, v. 3, n. 01, p. 154–162, 1995.