POLYPUBLIE

Goal-Oriented Calibration and Validation for Predictive Modeling

Antonin Paquette-Rufiange

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

Date	2024
Type	Mémoire ou thèse / Dissertation or Thesis Thèse de doctorat/Ph.D. thesis
URL de PolyPublie PolyPublie URL	https://publications.polymtl.ca/59208/
Conditions d'utilisation Terms of use	© Tous droits réservés / All rights reserved
Référence Citation	Paquette-Rufiange, A. (2024). Goal-Oriented Calibration and Validation for Predictive Modeling [Thèse de doctorat, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/59208/

Détails de la thèse
Thesis details

Directeurs de recherche Advisors	Marc Laforest, & Serge Prudhomme
Programme Program	Doctorat en mathématiques

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal
https://publications.polymtl.ca

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Goal-Oriented Calibration and Validation for Predictive Modeling

ANTONIN PAQUETTE-RUFIANGE
Département de mathématiques et de génie industriel

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
Mathématiques de l'ingénieur

Août 2024

© Antonin Paquette-Rufiange, 2024.

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Cette thèse intitulée :

Goal-Oriented Calibration and Validation for Predictive Modeling

présentée par **Antonin PAQUETTE-RUFIANGE**
en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

James GOULET, président
Marc LAFOREST, membre et directeur de recherche
Serge PRUDHOMME, membre et codirecteur de recherche
Teresa PORTONE, membre
Danial FAGHIHI, membre
Olivier LE MAÎTRE, membre externe

DEDICATION

À Luce, Mathias et Mélodie

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my advisors Prof. Marc Laforest and Prof. Serge Prudhomme. Their constant support and insightful guidance throughout the years have allow me to grow as a researcher. I am especially thankful for their confidence and the freedom they have given me to pursue my own research questions.

I would like to thank my friends Damien Tageddine and Ziad Aldirany for their support and help during the PDE courses. I also want to acknowledge my colleague Alexis Montoison for its ever enthusiastic help regarding programming with the programming language Julia. I am very thankful for its works on the libraries NOMAD.jl, KNITRO.jl and on ADNLPModels.jl, which have been extensively used in this thesis.

I also want to thank the committee members, Prof. Danial Faghihi, Prof. James Goulet, Dr. Olivier Le Maître, and Dr. Teresa Portone, for accepting to evaluate the thesis and participate to my defense.

Finally, I want to wholeheartedly thank my partner Luce for her unfailing patience and unwavering support throughout the years.

RÉSUMÉ

Nous proposons de nouvelles méthodologies afin d'améliorer les processsus de calibration et de validation de modèles physiques pour la prédiction de quantités d'intérêt. L'objectif général de la thèse est de proposer des approches qui prennent explicitement en compte les objectifs de prédiction d'un modèle, à savoir la prédiction de quantités d'intérêts sur les scénarios de prédiction.

De nombreuses approches ont été proposées afin de valider des modèles. La plupart de ces méthodes reposent sur la comparaison entre des données expérimentales et des prédictions issues du modèle, lesquelles sont obtenues à des scénarios de validation. Le choix de ces scénarios de validation sur l'évaluation de la validité du modèle est crucial, puisqu'un modèle peut être valide pour certains scénarios, mais invalide pour d'autres. La première contribution de cette thèse est le développement d'une méthodologie permettant le design d'expériences de validation. L'originalité de notre approche est que l'expérience de validation est conçue afin d'être la plus représentative pour valider la prédiction d'une quantité d'intérêt à un scénario de prédiction. La méthodologie repose sur le calcul de sensibilités obtenues par la méthode Active Subspaces.

Par conséquent, le processus de validation souhaite en général déterminer si les disparités entre les prédictions du modèle et les données expérimentales sont trop importantes. Le processus de calibration quant à lui a pour objectif de calculer la valeur des paramètres du modèle permettant de minimiser une certaine fonctionnelle des disparités entre le modèle et les données expérimentales. La deuxième contribution de cette thèse consiste à adapter l'approche de calibration de Kennedy et O'Hagan, où les erreurs de modélisation sont explicitement prises en compte. Notre approche présente deux nouveautés. D'une part, chaque valeur de paramètres de modèles génère ses propres erreurs de modélisation. D'autre part, les paramètres du modèle sont calibrés de sorte qu'une fonctionnelle des erreurs de modélisation est minimisée. Cette fonctionnelle peut alors être définie en fonction des objectifs de prédiction du modèle.

ABSTRACT

We propose in this work some goal-oriented methodologies for the validation and calibration of physical models. The objective of the thesis is to tailor these processes so that they a priori and explicitly take into account the predictive objective of a model, namely the prediction of quantities of interest at given prediction scenarios.

The validation process consists in comparing model predictions against experimental data in order to assess if the model can be employed for predictive purposes. Numerous validation approaches have been proposed, most of which suppose the comparison of experimental data and model predictions at given validation scenarios. The choice of the validation scenarios is important, as a model may be valid for certain scenarios, but invalid for others. The first contribution of this thesis is the development of a methodology to design optimal validation experiments. The novelty of our approach is that the optimal validation experiment is designed to be the most representative with respect to the task of validating the prediction of quantities of interest at a given prediction scenario. Our methodology employs the sensitivity method Active Subspaces in order to relate the validation experiment to the quantities of interest at the prediction scenario.

On one hand, the validation process determines whether discrepancies between the model predictions and experimental data are too important. On the other hand, the objective of the calibration process is to compute the value of the model parameters such that a certain functional of these discrepancies is minimized. The second contribution of this thesis consists in the adaptation of the calibration framework proposed by Kennedy and O'Hagan, where an explicit representation of the modeling discrepancy is introduced. Our calibration method presents two important novelties. First, each value of the model parameters generates its own modeling discrepancy. Second, the calibration of the model parameters is performed by minimizing a functional of the modeling discrepancy. The definition of this functional can be tailored toward the predictive objective of the model.

TABLE OF CONTENTS

DEDICATION iii

ACKNOWLEDGEMENTS iv

RÉSUMÉ v

ABSTRACT vi

TABLE OF CONTENTS vii

LIST OF TABLES x

LIST OF FIGURES xi

CHAPTER 1 INTRODUCTION 1
1.1 The Scientific Act of Modeling 1
1.2 Literature Review 3
1.3 Objectives of the Thesis 8
1.4 Outline of the Dissertation 11

CHAPTER 2 TOOLBOX OF THE VVUQ PRACTITIONER 12
2.1 Preliminary Notions and Notations for Predictive Modeling 12
2.2 Sampling Methods 14
2.3 Sensitivity Analysis 17
2.4 Gaussian Processes 24
2.5 Brief Overview of Optimization Methods 27
2.6 Inference/Calibration Methods 32
2.7 Modeling Errors 37
2.8 Validation 41

CHAPTER 3 OPTIMAL DESIGN OF VALIDATION EXPERIMENTS FOR THE PREDICTION OF QUANTITIES OF INTEREST 45
3.1 Introduction 45
3.2 Terminology and Preliminary Notions about Validation and Predictive Modeling 49
3.2.1 Abstract Model 49
3.2.2 Classes of Parameters 50

3.2.3 Classes of Scenarios . . . 51
3.2.4 Observables and Quantities of Interest . . . 52
3.2.5 Uncertainties and Modeling Errors . . . 54
3.2.6 Calibration Process and Optimal Design of Calibration Experiments . 56
3.2.7 Validation Process . . . 59
3.3 Optimal Design of Validation Experiments . . . 61
3.3.1 Sensitivity Analysis . . . 61
3.3.2 Active Subspace Method . . . 62
3.3.3 Optimal Choice of the Validation Scenario . . . 63
3.3.4 Optimal Choice of Observables and Sensor Parameters . . . 64
3.4 Numerical Examples . . . 65
3.4.1 The Projectile Problem . . . 66
3.4.2 Contaminant Transport Problem . . . 72
3.5 Conclusion . . . 78
3.A Velocity Field for the Pollutant Transport Example . . . 80
3.B Gradient for the Contaminant Transport Example . . . 81
3.C Solution of the Optimal Design Problems . . . 84

CHAPTER 4 GOAL-ORIENTED CALIBRATION OF A MODEL AND ITS MODELING ERRORS . . . 87
4.1 Introduction . . . 87
4.2 Preliminary Notions about Predictive Modeling and Modeling Errors . . . 91
4.2.1 Illustrative Model . . . 91
4.2.2 Abstract Model of the System of Interest . . . 92
4.2.3 Experiments on the System of Interest . . . 93
4.2.4 Modeling Errors . . . 96
4.2.5 Another Representation of the Modeling Errors . . . 99
4.3 Calibration of Model Parameters and Modeling Errors . . . 101
4.3.1 Calibration in the Absence of Modeling Errors . . . 101
4.3.2 Calibration in the Presence of Modeling Errors . . . 104
4.3.3 Calibration Objective vs Predictive Objective . . . 106
4.4 Goal-Oriented Calibration . . . 108
4.4.1 Calibration of the Modeling Discrepancy . . . 108
4.4.2 Definition of the Goal-Oriented Calibration Problem . . . 110
4.5 Numerical Examples . . . 112
4.5.1 Synthetic Example . . . 113

4.5.2 Verification Example 122
4.5.3 Sandia Validation Challenge - Thermal Problem 125
4.6 Conclusion 132

CHAPTER 5 CONCLUSION 135
5.1 Summary of Works 135
5.2 Limitations and Perspectives 137

REFERENCES 139

LIST OF TABLES

Table 3.1 Summary of the parameters, observables, and QoI as well as their source of uncertainty 57

Table 3.2 Prediction scenario x_{pred}, controlled environment $\mathcal{X}_{\text{lab}}$, and tentative validation scenario $x_{\text{val,tent}}$ 67

Table 3.3 Description of the prediction scenario x_{pred} and of the controlled environment $\mathcal{X}_{\text{lab}}$ for the pollutant transport model. 75

Table 3.4 Results of the ten verification experiments for the projectile problem 86

Table 3.5 Results of the ten verification experiments for the pollutant transport problem 86

Table 4.1 Subset of the data from the thermal problem. Taken from Table 1 in [1] 94

Table 4.2 Choice of the norm for the functional J 116

Table 4.3 Choice of the norm for the functional J 124

Table 4.4 Optimal model parameters θ^* for different calibration objectives . . . 125

Table 4.5 Optimal modeling discrepancy parameters $\xi^*(\theta^*)$ for different calibration objectives 125

Table 4.6 Experimental design for the calibration and accreditation data 126

Table 4.7 $\text{RMSE}_{\text{calib}}$, loglikelihood, and functional of the quantile error function for the calibration approaches 128

Table 4.8 $\text{RMSE}_{\text{accred}}$ for the calibration approaches 132

LIST OF FIGURES

Figure 1.1 Extrapolation of the validation process from the observable to the QoI. The x-axis represent a "distance" between the validation and prediction scenario. The y-axis represent a "distance" between the observable functional and the QoI functional. 9

Figure 2.1 Monte Carlo sample (left) and Latin Hypercube sample (right) for $N = 20$. 16

Figure 2.2 Function (2.23) for $n = 1$ (left) and for $n = 4$ (right). 23

Figure 2.3 Illustration of Gaussian processes. The dashed blue lines indicate the mean function whereas the shaded areas represent the 95% prediction interval. Three realizations of the Gaussian processes are displayed by thin lines. 25

Figure 2.4 Illustration of non-stationary Gaussian processes constructed from the Gaussian kernel. The dashed blue lines indicate the mean function whereas the shaded areas represent the 95% prediction interval. Three realizations of the Gaussian processes are displayed by thin lines. . . 26

Figure 2.5 Illustration of two poll steps of the MADS algorithm. The two frames are depicted by thick squares and the candidate points are shown as points. If no candidate points decrease the objective function (left), the frame and mesh are reduced and new candidate points are considered (right). 31

Figure 3.1 Illustration of the various scenarios. The prediction scenario may be producible (black circle) or may not be producible experimentally (black square). The validation and calibration scenarios are elements of the set $\mathcal{X}_{\text{lab}}$ of controlled environments. 52

Figure 3.2 Diagram of the interactions between the various parameters and the system of interest as well as the model. It is important to observe that the model parameters θ do not affect the system of interest, but only the observation and QoI functional h_{obs} and h_{qoi}. For the sake of simplicity, the QoI functional h_{qoi} does not depend explicitly on the sensor parameters z. 53

Figure 3.3 Eigenvalues (top row) and eigenvectors (bottom row) associated with the influence matrices of the QoI (3.15) for the prediction scenario x_{pred} (left), the tentative validation scenario $x_{\text{val,tent}}$ (center), and the optimal validation scenario x_{val} (right). For the sake of clarity, only the first four eigenvectors associated with the largest four eigenvalues are displayed as the values of the other two eigenvalues are negligible. 68

Figure 3.4 Objective functional of Problem (3.14) associated with the observation functionals $h_{\text{obs}}(x_{\text{val}}, \Theta, z) \coloneqq u$ and $h_{\text{obs}}(x_{\text{val}}, \Theta, z) \coloneqq \mathrm{d}^2u/\mathrm{d}t^2$. 69

Figure 3.5 Position u (top row) and discrepancies (bottom row) for the prediction scenario (left column), tentative validation scenario (center column), and the optimal validation scenario (right column). The experimental observations are in red and the model observations are in blue. The shaded areas represent the 95% confidence interval for both the experimental and model observations. The absolute value of the discrepancies (3.5) are computed at the sensor location indicated by the black dashed lines. 72

Figure 3.6 Domain Ω and its boundaries for the contaminant transport model. The regions for which we seek the mean pollutant concentration are indicated by Ω_1 and Ω_2 (respectively employed for the definition of the QoIs). The grid is used to indicate the regions Ω_{obs} in which the mean concentration of pollutant can be observed. 74

Figure 3.7 Concentration of the pollutant for the prediction and optimal validation scenarios: (top) prediction scenario, (center) optimal validation scenario associated with $h_{\text{qoi},1}$, (bottom) optimal validation scenario associated with $h_{\text{qoi},2}$. 76

Figure 3.8 Eigenvalues (top row) and eigenvectors (bottom row) associated with the influence matrices for the prediction scenario (black circles) and for the validation scenario (red squares). The left and right columns display the sensitivity indices related to the first and second QoIs, respectively. For the sake of clarity, only the first (and only relevant) eigenvector is displayed for each influence matrix. 77

Figure 3.9 Value of the objective function in (3.14) for the first QoI (top) and second QoI (bottom). 78

Figure 3.10 Streamline of the velocity field v 81

Figure 3.11 Mesh employed to compute the optimal validation scenario for the prediction of the mean concentration inside Ω_1. 85

Figure 4.1 Illustration of the Gaussian process $Y^{\theta,\xi}$ (top row) and the modeling discrepancy E^{ξ} for the thermal problem. The experimental data and the corresponding residuals are displayed by various marks. The mean functions of the Gaussian processes are displayed in dashed and some realizations are shown with thin lines. 99

Figure 4.2 Illustration of the quantile error function ε_E from the modeling discrepancy E (illustrated in blue). The dashed curve indicates the mean function of the Gaussian process while the shaded area represent the 95% prediction interval. Two sets of residuals are displayed for illustrative purpose. 112

Figure 4.3 Experimental data and model prediction for $\theta = 0.45$. 113

Figure 4.4 (Left) Optimal representation of the modeling discrepancy E obtained by solving Problem (4.30). (Right) Optimal representation of the modeling discrepancy E obtained by sequentially solving problems (4.32) and (4.34). Both representation of the modeling discrepancy are obtained for $\theta = 0.449$. The dashed blue line is the mean of the Gaussian process while the blue-shaded area is the 95% prediction interval. Some realizations of the priors are displayed by thin lines. 115

Figure 4.5 Quantile error function (4.38) for the modeling discrepancy obtained with the first and second calibration approaches. The value of α is fixed at 95%. 116

Figure 4.6 Profiles of the calibration functional for the first objective (left), the second objective (middle), and the third objective (right). 118

Figure 4.7 Goal-oriented calibration for the first objective (left), the second objective (middle), and the third objective (right). The top row illustrates the optimal model predictions. The middle row presents the associated residuals and modeling discrepancy $E^{\xi^*(\theta^*)}$. The bottom row displays the corresponding quantile error functions. 119

Figure 4.8 Profiles of the calibration functional for the first objective (left), the second objective (middle), and the third objective (right) with the representation (4.43) of the modeling discrepancy. 120

Figure 4.9 One possible optimal solution to the calibration Problem (4.25). The optimal model prediction and associated modeling discrepancy are displayed for the optimal model parameters $\theta^* = 0.3$. 123

Figure 4.10 Surface of response of the manufactured solution and the design of experiments marked by black crosses. 123

Figure 4.11 Optimal model predictions and associated modeling discrepancy at the four calibration scenarios. The first and third rows show the predictions obtained with the various calibration methods. The second and fourth rows depict the modeling discrepancies and residuals obtained with the second objective of the goal-oriented calibration. Some realizations of the modeling discrepancies $E^{\xi^*(\theta^*)}$ are also demonstrated by thin curves. 130

Figure 4.12 Accreditation data and model predictions for the second goal-oriented calibrated model (black), the MLE calibrated model (red) and the MLE KOH calibrated model (violet). The Gaussian process Y^{θ^*} for the second goal-oriented calibration is displayed in green. 132

CHAPTER 1 INTRODUCTION

1.1 The Scientific Act of Modeling

The modeling process is a central activity of any scientific activity. From observations about a system of interest, one formulates hypotheses and simplifications regarding the underlying phenomenon. The conceptualization and abstraction of a system of interest allows one to build a model that relates parameters to the observables. The model is thus fundamental for understanding the system of interest, as it rationalizes and renders intelligible the realm of reality. Moreover, a model may be employed to predict quantities of interest (QoIs) pertaining to the system of interest, quantities that may not be accessible otherwise. For example, one may want to predict the behavior of a system in the future. It is only through the model that one can extrapolate the future states of the system of interest. Another example of a QoI accessible only via a model could be the maximal stress in a solid. We can record strain at some regions of the domain, but the location and value of the maximal stress cannot be directly recorded.

Since the 19th century, there has been an exponential growth in scientific knowledge spanning ever increasing fields of applications. Combined with substantial developments of the mathematical language and computational ability, we are now able to describe highly complex phenomena with detailed models. Such models often involve ordinary or partial differential equations (ODEs and PDEs), large linear and/or non-linear systems of equations, and elaborated optimization problems. Finally, the recent advances in Machine Learning (ML) and the profusion of scientific data has given rise to Scientific Machine Learning (SciML) which aims at creating effective semi-empirical models trained on data, without deep prior knowledge.

In general, the various hypotheses made during the modeling process introduce unknown parameters, the so-called model parameters. They could represent a variety of quantities: They may relate to material properties, regression coefficients, weights and biases in a neural network, and so on. Since they are not precisely known, we need a procedure to determine their value, which is exactly the goal of the calibration. Roughly speaking, any calibration method compares the model predictions with experimental data gathered from the system of interest to infer the value of the model parameters.

However complex these models are or are destined to become, they remain and will only remain mere abstractions of a system of interest. They rely on hypotheses, assumptions, and simplifications, which in turn generate modeling errors. These modeling errors may stem

from influential contributions that may have been neglected as input parameters. Modeling errors are also created by the structure of the model since it cannot encompass all possible phenomena influencing the system of interest. An additional and subtle source of modeling errors is the actual value of the model parameters. A poorly calibrated model would not yield predictions close to the observations. It is thus the whole objective of the modeler to capture the important features of the system of interest and to calibrate adequately the model parameters. By carefully pondering the conception and the calibration of a model, the modeling errors can be reduced, but some will inevitably remain.

The process to assess whether or not a model is a reliable representation of the system of interest is called validation. In a nutshell, experimental data are compared to the model predictions to estimate the modeling errors and the mismatch is thus quantified and aggregated. If the modeling errors due to the model are deemed acceptable, then the model is considered valid (not invalid). Otherwise, the model is deemed invalid. An important aspect of the validation process is that we want to validate a model for its intended use. For example, a linear elastic model may be useful to describe small deformations of a linear elastic material, but this same model would be less useful to describe plastic deformations.

The main objective of this thesis is to extend and improve current validation and calibration processes in such a way that they are specifically tailored toward the predictive objective of the model. Indeed, the current state of the literature studying the validation problem scarcely addresses the problem of validating a model with respect to a non-observable QoI. Answering this question is deemed important, since the objective of calculating model predictions often consists in computing a non-observable QoI. In other words, we want to know if the model is valid to perform predictions of a QoI, not necessarily if the model is valid to predict observables. It is well-known from prior studies on goal-oriented estimation of discretization errors, that these two different objectives can produce significantly different solutions, both in terms of accuracy and cost. The first contribution of this thesis consists in the development of a novel approach to design optimal validation experiment. These experiments are coined optimal because they are specifically constructed to be representative (in some sense) of the QoI to be predicted.

Furthermore, the literature often introduces the calibration problem as the process that approximates the true value of the model parameters. This point of view implicitly assumes that the model is a fairly good representation of the system of interest, i.e., the modeling errors are negligible. However, this notion of true model parameters is not always well-founded, since we do not know a priori if the modeling errors are small. Moreover, this perspective confines the scope of the calibration problem to that of only seeking approximations of

true values. Hence, the second contribution of this thesis aims at proposing a different paradigm, that of shifting the calibration objective by acknowledging that we do not seek the true model parameters, but rather optimal ones. The proposed calibration approach directly approximates the modeling errors. We then formulate the calibration objective as the minimization of a functional that estimates the modeling errors. In this way, the calibration process can be tailored in such as way that specific features of the modeling errors are minimized, features that may be relevant toward the predictive objective.

1.2 Literature Review

The literature on Verification, Validation, and Uncertainty Quantification (VVUQ) is vast and spans several scientific fields. We will narrow our review of the literature to four different VVUQ activities. It is not the objective here to dwell into the details, but to provide an overview of the currently available methodologies. Certain approaches, directly relevant to the thesis will be described more thoroughly in Chapter 2. For a general overview of the VVUQ challenges and state-of-the-art, we refer the reader to the books [2, 3].

The first VVUQ activity focuses on goal-oriented error estimation. Altought this subject is not directly studied in this thesis, the discussion will provide context regarding the challenge of adapting and extending the VVUQ activities to the prediction of QoI. The second VVUQ subject, studied more extensively in Chapter 4, concerns the calibration problem. The third subject is dedicated to the modeling errors. The last subject of the literature review is the validation problem.

Verification - Goal-Oriented Error Analysis

The verification problem consists in estimating and assessing errors in approximations of the mathematical model describing the system of interest. It is very rarely the case that an exact analytical solution can be computed from the mathematical model considered. We must resort to approximations, e.g. via discretization methods such as finite differences, finite elements and time integrators to solve differential equations. It is then important to assess the discretization errors associated to the approximated solution. Several methods and approaches have been developed in the case of the finite element method, see e.g. [4, 5]. The QoI rarely describes the whole solution of the mathematical model, but rather consists of some specific features of the solution. Goal-oriented error estimation techniques have been developed to specifically assess the discretization errors pertaining to QoIs, see e.g. [6, 7] and references therein. These approaches are essentially a posteriori, in that the estimation of the

discretization errors in the QoI is performed after a solution is computed. We nevertheless mention the work of Kergrene et al. [8] that differs from the previous works in that the prediction of the QoI is encoded a priori in the mathematical model to be solved. The adaptation of the calibration and validation processes presented in this thesis share the same idea of incorporating a priori the predictive goal.

As we shall see, the literature regarding the goal-oriented calibration and validation of a model is much scarcer than goal-oriented verification. We tentatively explain this fact by two reasons. First and foremost, the verification activity comes first in the whole process of determining the accuracy of simulations predictions. Indeed, it is useless (and somewhat dubious) to calibrate and validate a model with respect to a QoI if the discretization and truncation errors are not well controlled and estimated. The second reason, which motivates parts of this thesis, is that the calibration and validation problems do not possess unique and unambiguous definitions compared to the verification problem: the definition of the verification problem is clear, one wishes to estimate the numerical error in the approximate solution to a mathematical problem. Goal-oriented verification adds the supplemental aspect of estimating the errors on the approximated QoI. On the other hand, the calibration and validation problems are not uniquely defined, so the goal-oriented extension of these activities is not straightforward.

Calibration

The calibration problem, along with the validation problem, are pervasive to any scientific study. Indeed, a model often involves quantities that must be inferred, the model parameters, in order to perform predictions. The literature addressing this problem is rather vast as each model and application present specific challenges. For example, one could want to infer finite-dimensional model parameters such as the Young modulus and Poisson ratio of a solid under external loads. One could also desire to infer the infinite-dimensional permeability of a soil to predict contaminant flows. One could also want to infer the states of non-stationary phenomena, where noisy observations are performed at finite time values. These three examples illustrate the vast scope of the calibration problem. We invite the interested reader to consult the books [3, 9–11] for a more thorough presentation of the various calibration problems and the methodologies adopted to solve them. In this thesis, we will focus only on the first type of calibration, namely the inference of finite-dimensional model parameters. We will present general references on the ubiquitous Maximum Likelihood Estimation (MLE) and Bayesian approach. We will then review several works that perform calibration by taking into account the fact that the model is an imperfect description of the reality.

The theory of the Maximum Likelihood Estimation is extensively covered in the books [12,13]. Notably, the asymptotic consistency and the asymptotic normality of the MLE estimator are discussed in these references. The paper by White [14] discusses the properties of the MLE estimator for a misspecified model, that is, a model that is not a perfect representation of the system of interest.

The Bayesian approach to calibration is presented in the books [9, 10, 12, 13]. As with the MLE estimator, the references [12,13] present asymptotic consistency of estimators based on the posterior as well as the asymptotic normality of the latter. The paper [15] adapts the results in the case where the model is misspecified. The Bayesian approach to calibration in the infinite dimensional case is discussed in [3, Chapter 6], [16] and [17]. These works discuss the well-posedness of the posterior, namely the existence of a unique posterior, which is stable with respect to perturbations in the data.

The presence of modeling errors affects the calibration process. This recognition has resulted in the development of various calibration approaches. For example, the seminal work of Kennedy and O'Hagan [18] (KOH) proposes to describe the modeling discrepancy on the observables by a Gaussian process. The parameters of this Gaussian process need to be calibrated alongside the model parameters. Hence, the calibration problem now aims at simultaneously computing the value of the model parameters as well as the value of the modeling discrepancy parameters. Brynjarsdóttir and O'Hagan [19] further study the KOH framework by providing prior information regarding the modeling discrepancy by conditioning the Gaussian process. Jiang et al. [20] present a review of the KOH calibration as well as optimization-based calibration. They provide some definitions for optimal model parameters and discuss the a posteriori computation of the modeling discrepancy. Moreover, they propose a novel calibration approach that sequentially ponders whether the posterior should be updated. Leoni et al. [21] adopt the KOH framework, but acknowledge the different nature of the model parameters and modeling discrepancy. Each value of model parameters produces its own model, with its own modeling errors. Given some experimental data and for each value of the model parameters, the authors compute an optimal representation of the modeling discrepancy.

The work of Hu et al [22] calibrates the model parameters by minimizing the area validation metric (which will be described in Section 2.8). Their methodology takes into account both aleatory and epistemic uncertainties. The paper by Joseph and Melkote [23] formulates the calibration problem as a Least Squares problem. In the case that their calibrated model is not deemed valid, they sequentially add regression terms to the model output until the model is considered valid. He and Xiu [24] propose to correct the model parameters and

the model outputs with polynomial functions while also enforcing some physical constraints. The calibration problem is cast as a residual minimisation problem with constraints where the polynomial corrections are sought.

Modeling Errors

Since no model provides a perfect abstraction of reality, every model produces modeling errors, more or less consequential. We present here some methodologies that explicitly approximate modeling errors.

As mentioned previously, the Kennedy and O'Hagan approach (KOH) allows the simultaneous calibration of the model parameters and the modeling discrepancy. As a result, an approximation of the modeling discrepancy affecting the observables is obtained across the calibration scenarios. A limitation of this approach is that the modeling discrepancy on the observables cannot be extrapolated to those in non-observable QoIs.

A way to circumvent this issue have been proposed by Sargsyan et al. [25]. Instead of viewing the model parameters as a deterministic quantity to be estimated, they consider the model parameters as aleatory. More precisely, each experiment on the system of interest corresponds to a realization of the model parameters. This way, some variability in the experimental data can be captured by the structure of the model. However, the contributions to the modeling errors that cannot be described within the structure of the model cannot be represented with this approach. Nevertheless, the modeling errors encoded in the model parameters can now be straightforwardly propagated to the QoIs.

Oliver et al. [26] recognize the fact that modeling errors arise from limitations of the model. They propose to directly enrich the equilibrium equations of the model with stochastic terms in order to obtain an augmented model. The additional stochastic terms should be rich enough as to capture phenomena that could not be previously represented by the model. This approach is different from that proposed by Kennedy and O'Hagan because the additional stochastic terms affect the equilibrium relations of the model, not merely the observables function. The introduction of these new stochastic terms usually add new model parameters to be calibrated. Using this approach, the propagation of the modeling errors to the QoIs can also be performed, because an approximation of the modeling errors is now encoded in the structure of the augmented model. A major limitation of this approach is the fact that it is intrusive, i.e., modification to the original solver must be made in order to solve the equilibrium relations of the augmented model.

Validation

The validation problem is ubiquitous to all scientific domains and applications. In order to provide relevant and accurate information about a system of interest, a model should be validated, or more precisely, not invalidated. In the following presentation, we will focus on works that propose a generic methodological approach to the validation problem, rather than discussing solutions specific to a scientific domain. Again, we do not aim here to be exhaustive, but rather to provide some references that influence the research done in this thesis.

We first mention the works of Oberkampf et al. [27,28] that provide a comprehensive overview of the verification and validation problems. The authors also provide an extensive review of the literature. Notably, they discuss the difficulties to validate a model with experimental data that do not cover all the possible prediction scenarios. The paper of Roy and Oberkampf [29] discusses in detail the various sources of uncertainties and errors that may affect the quality of a prediction. They adopt different approaches to propagate the uncertainty based on whether they are considered aleatory or epistemic.

One way to perform the validation assessment is via the use of validation metrics. Ferson et al. [30] propose the area-validation metric which computes the integral of the difference between the empirical distribution and the distribution of the observables predicted by the model. Rebba and Mahadevan [31] propose the model reliability metric which computes the probability that the difference between the experimental data and model predictions is smaller than a given threshold. The paper by Mullins et al. [32] discusses model reliability when multiple validation scenarios are considered.

Instead of validating a single model, one can seek to perform model selection, where multiple models are compared. This can be performed with Bayes factor, where the model evidence of two competing models is compared [33,34]. The notion of posterior model plausibility [35–37], a normalized version of the model evidence, can also be used to rank the most probable model, among a set of candidate models, given the experimental data. A systematic approach to model calibration, selection, and validation is the Occam-Plausibility Algorithm [36–38] that aims at identifying the simplest not invalidated model.

Oden and Prudhomme [39] propose to validate a model with respect to the prediction of QoIs by performing two calibration steps. More precisely, a first calibration is realized with calibration data, then another calibration is done with calibration and validation data. If the discrepancies between the two models are too large the model is deemed invalid. Morrison et al. [40] extend this idea in the context of cross-validation. Several partitions of the data

are considered, where each partition consists in a calibration and a validation data sets. For each such partition, the model is calibrated and its distribution of QoIs is compared to that of the model calibrated with all the data. This procedure allows the identification of the partition that generates the worst discrepancy in the observables and the QoIs, which then can be used to assess the validity of the predicted QoI.

1.3 Objectives of the Thesis

The main objective of this thesis is to adapt and extend existing validation and calibration methodologies so that they are specifically tailored toward the predictive objective of the model. The following three postulates constitute the backdrop that has mainly influenced this research work.

- No model is a perfect representation of reality. Inevitably, there are some modeling errors associated with any model.
- Reality is only accessible via the gathering of experimental data. Importantly, these experimental data consist only in observables.
- The predictive objective of a model is the prediction of QoIs that may or may not be observable or observed.

These postulates are clearly intertwined and affect both the validation and calibration processes. The latent objective of the thesis is to untangle all these ideas in the context of calibrating and validating a model. Since the objectives and scope of the validation and calibration are very different, we break up the overall objective into two sub-objectives.

The first sub-objective consists in adapting the validation process so that we can assess whether a model is valid for the prediction of non-observable QoIs. The challenge is that we cannot acquire experimental data for non-observable QoIs, i.e., we cannot interrogate the system of interest for these QoIs. Figure 1.1 describes two manners by which a QoI can be non-observable and how this issue could potentially affects the validation assessment.

The first reason a QoI may be non-observable is because the prediction scenario is not producible experimentally, e.g. future predictions or predictions in complex/extreme operational conditions. The "distance" between the validation and prediction scenarios is depicted by the x-axis in Figure 1.1. The second reason a QoI may be non-observable is because it represents a quantity not directly measurable from the system of interest, e.g. minimal/maximal/mean quantities. The "distance" between the observable functional used for the validation and the

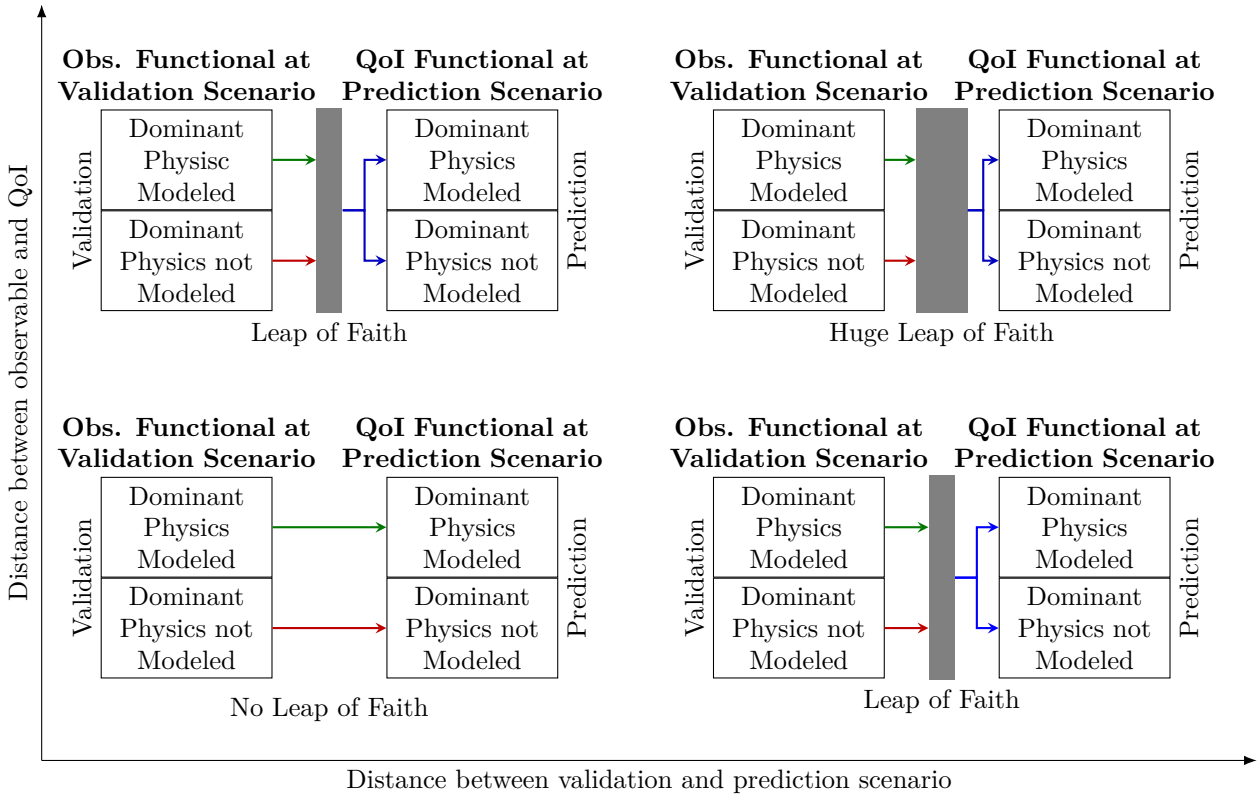


Figure 1.1 Extrapolation of the validation process from the observable to the QoI. The x-axis represent a "distance" between the validation and prediction scenario. The y-axis represent a "distance" between the observable functional and the QoI functional.

QoI functional is depicted by the y-axis in Figure 1.1. In the specific case where the QoI is observable and the prediction scenario can be producible experimentally, then the validation process directly assesses whether or not the model can be employed. However, as soon as the prediction scenario is not producible experimentally or the QoI cannot be abstracted as an observable functional, there is a leap of faith between the validation assessment and the prediction. More precisely, we now need to extrapolate the result of the validation onto the prediction of the QoI. This extrapolation gap does depend upon the "distance" to the origin along the y-axis. The whole challenge of this first sub-objective consists in defining this notion of "distance" and to adapt the validation process as to reduce as much as possible the leap of faith.

The second sub-objective of this thesis focuses on tailoring the calibration process toward the predictive goal of the model. More precisely, we will consider that the predictive objective when using a model is to predict some non-observable QoIs. Importantly, the predictive objective does not consist in inferring the values of the model parameters. This sub-objective implicitly involves the need to change the common viewpoint about calibration: since no model is perfect, the notion of true model parameters is not well defined. This a priori simple assessment nevertheless complicates our view of the calibration process in two closely related ways:

1. The calibration process should then be seen as the process that computes an optimal value for the model parameters;

2. Whatever the value given to the model parameters, there will always remain some modeling errors associated with the given model.

The challenge associated to this sub-objective is to define a calibration process that simultaneously takes into account the modeling errors and the predictive objective when using the model. In doing so, the procedure should calibrate both the model parameters and the modeling errors with the predictive objective in mind.

In this thesis, we do not study the calibration and validation problems in the context of a model described by multiple sub-models. The presence of sub-models or subsystems raises the following questions regarding the calibration and the validation problems:

- How should one calibrate the model parameters of the sub-models? Does the predictive objective of the *whole* system should be taken into account in these calibration problems?

- How does the validity assessment of the sub-models inform the validity of the *whole* system? What type of validation experiments should be performed on the subsystems to infer the validity of the prediction of QoI for the *whole* system?

For the sake of clarity and to narrow the focus of our work, the models will be considered stand-alone, in other words, not subjected to sub-models.

1.4 Outline of the Dissertation

The thesis is structured as follows. Chapter 2 presents an overview of methods employed throughout the thesis. The objective is to perform a unified and coherent presentation of the calibration, validation, and sensitivity analysis problems. We will also complement this presentation with a discussion about sampling methods and Gaussian processes. Moreover, a very brief account on the optimization methods employed will be made. Chapter 3 is essentially based on the article [41]. In this work, we present the methodology that was developed to design optimal validation experiments with respect to the prediction of a QoI. With the Active Subspace method, we define influence matrices that are employed to quantify the influence of the various parameters on the model predictions and QoIs. The design of optimal validation experiments is then obtained by sequentially solving two optimization problems. The first one defines the validation scenario to be carried out, while the second one specifies the type of measurements to be performed. The methodology is tested on two synthetic problems. In Chapter 4, we propose our novel approach on the calibration problem. We revisit the framework laid out by Kennedy and O'Hagan [18] to discuss the calibration of model parameters with and without the explicit presence of modeling discrepancy. Afterwards, the goal-oriented calibration approach is designed as two nested calibration problems. For each value of the model parameters, we first calibrate the modeling discrepancy so as to better fit the experimental data. Using the calibrated modeling discrepancy, we then evaluate a calibration functional for the given value of the model parameters. We then seek the value of the model parameters that minimizes this calibration functional. We illustrate the goal-oriented calibration approach on two synthetic problems and on the Thermal Problem of the Sandia Validation Challenge [1, 42]. Chapter 5 summarizes the thesis by recalling the key contributions. Finally, several new research avenues are suggested.

CHAPTER 2 TOOLBOX OF THE VVUQ PRACTITIONER

The objective of this chapter is to present an overview of some of the essential methods and techniques employed in Chapter 3 and in Chapter 4. Brief descriptions of these methods were provided in each article, but a unified and more precise presentation would provide valuable context for the thesis. Indeed, a significant amount of research has been dedicated to the study of these methods. We begin this chapter with some general notions and notations that will be employed throughout the thesis. Then, we proceed with the various subjects by attempting as much as possible to arrange them in logical order. We have not found any single source that would completely cover the material, so references will be provided when appropriate.

2.1 Preliminary Notions and Notations for Predictive Modeling

A deterministic model represents an abstraction of the system of interest. A deterministic model is described by a set of equations

$$r(x, \theta, u) = 0 \tag{2.1}$$

relating the input parameters $x \in \mathcal{X}$, the model parameters $\theta \in \mathcal{T}$, and the state $u \in \mathcal{U}$. We will assume that $\mathcal{X}$ and $\mathcal{T}$ are of finite dimension and that $\mathcal{U}$ is a Banach or a Hilbert space. Moreover, we assume that the equilibrium relation (2.1) is well posed, i.e., for each tuple $(x, \theta) \in \mathcal{X} \times \mathcal{T}$, there exists a unique and stable solution $u \in \mathcal{U}$ to (2.1). We will denote by g the function $g(x, \theta) = u$.

The model parameters θ represent the parameters that arise from the simplification and hypotheses inherent to the model. For example, they may describe material properties, regression coefficients, dimensionless quantities, and so on. They are usually not well-known, which is why we want to calibrate them against experimental data. Sometimes, the rationales and hypotheses from which the model parameters θ arise may not hold throughout $\mathcal{X}$. There will be slight differences in the meaning of the input parameters x between Chapter 3 and Chapter 4, but these will be addressed when appropriate. For this chapter, it is only necessary to view them as the parameters that describe the conditions at which we simultaneously subject the system of interest and the model.

Very often, we are not interested in the whole state solution u, but rather on a value $y \in \mathbb{R}$

given by the functional h

$$y = h(x, \theta, u) = h(x, \theta, g(x, \theta)) = h(x, \theta), \tag{2.2}$$

where we slightly abuse the notation to simplify the presentation. Depending on the regularity of the state u, h may represent pointwise evaluation of u and/or its derivative. It may also describe mean/minimal/maximal quantities.

Throughout the thesis, we will describe uncertainty within the probabilistic framework. We refer the interested reader to [43] for other representations of uncertainty. We denote the probability space $(\Omega, \mathcal{F}, \mu)$, where Ω is the sample space, $\mathcal{F}$ is the σ-algebra describing the possible events, and μ a probability measure. For the sake of the presentation, let us consider that the input parameters are uncertain. We assimilate the input parameters as the random variable X, which is a measurable function $X : \Omega \to \mathcal{X}$. The distribution of X is defined as $\mu_X(B) = \mu(X^{-1}(B))$ for $B \in \mathcal{B}$, where $\mathcal{B}$ is a σ-algebra of $\mathcal{X}$. The cumulative distribution function of X will be noted F_X. Often, we will describe the random variable X with its density f_X, which will be the Radon-Nikodym derivative with respect to the Lebesgue measure. With the distribution or density, we can compute the first and second moments of X, denoted by $\mathbb{E}(X)$ and $\mathrm{Cov}(X)$, respectively.

By considering either the input parameters or the model parameters random, we generate a stochastic model. In this thesis, we will not seek to compute the solution $u : \Omega \to \mathcal{U}$ that takes into account the stochastic element. Since we only transform finite dimensional parameters into a finite number of random variables, the type of stochastic models considered in this work is relatively restricted:

- They do not involve Brownian noise as in stochastic differential equations [44];
- They do not involve parameters that are stochastic processes, e.g. material properties that vary inside a domain. In this sense, the methods developed to solve stochastic partial differential equations, such as the collocation method or the stochastic Galerkin approximation [45, 46] are not considered here.

We shall adopt here a Monte Carlo approach described in Section 2.2 to obtain samples of the solution u. As mentioned earlier, we are not necessarily interested in the whole solution u, but rather in one or multiple functionals h. By rendering x or θ random, we also make y in Equation (2.2) aleatory. Let us consider for example that the model parameters are random and noted Θ. The resulting random variable $Y(x)$ is thus defined by $Y(x, \omega) = h(x, \Theta(\omega))$. We will also denote the random variable $Y(x)$ by $h(x, \Theta)$. The distribution of $Y(x)$ is

obtained by propagating the uncertainty $\mu_{Y(x)}(B) = \mu(\Theta^{-1}(\tilde{Y}_x^{-1}(B)))$, where $B \in \mathcal{B}$ and $\mathcal{B}$ is the Borel σ-algebra of the real numbers. The function $\tilde{Y}_x$ is defined as

$$\begin{aligned} \tilde{Y}_x : \mathcal{T} &\to \mathbb{R} \\ \theta &\mapsto h(x, \theta). \end{aligned}$$

If a collection of random variables Θ_i is identically and independently distributed according to Θ, then the collection of random variables $Y_i = h(x, \Theta_i)$ are identically and independently according to $Y(x)$. As is customary in the literature, the random variables are written using an upper-case and realizations of their respective counterparts are denoted with a lower-case. Moreover, the underlying probability space $(\Omega, \mathcal{F}, \mu)$, as well as the explicit dependence of a random variable upon ω, will be omitted, unless relevant to the discussion.

2.2 Sampling Methods

Often, the analytical expression of the distribution or the density of a random variable is not tractable. For example, the distribution of the random variable $Y(x) = h(x, \Theta)$ is known for very restricted cases of $h(x, \cdot)$ and random variable Θ, e.g. linear h and Gaussian Θ. Also, the posterior in Bayesian analysis is analytically known in limited cases of likelihood and prior, such as conjugate prior. However, we are usually interested in specific features of these intractable random variables, such as their moments and their quantiles. In the thesis, we will focus solely on the problem of propagation of uncertainty via a function, such as in $Y(x) = h(x, \Theta)$. Two ubiquitous methods to perform this task is the Monte Carlo (MC) approach and the Latin Hypercube Sampling (LHS). The presentation of these two methods is heavily inspired by the unpublished book by Owen [47].

Monte Carlo Sampling

The primary purpose of Monte Carlo is to approximate the mean $\mathbb{E}(Y(x))$ by using the Strong Law of Large Numbers [12, Corollary 1.63] and [48, Theorem 10.12]. Let us consider a sequence of N random variables Θ_i independent and identically distributed (iid) according to Θ. In the thesis, we suppose that the d components of Θ are independent and that we can sample each component. The independence between the d components is not necessary here, but merely a hypothesis. From this sequence, we can construct another sequence $Y_i(x) = h(x, \Theta_i)$ which is independent and identically distributed (iid) according to $Y(x) = h(x, \Theta)$.

Then, the Strong Law of Large Number states that the sample mean

$$\mu_{MC} = \frac{1}{N}\sum_{i=1}^{N} Y_i(x) = \frac{1}{N}\sum_{i=1}^{N} h(x, \Theta_i) \tag{2.3}$$

converges almost surely to $\mathbb{E}(Y(x))$ as $N \to \infty$. The estimator μ_{MC} is unbiased and its variance is

$$\mathrm{Var}\left(\mu_{MC}\right) = \mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N} Y_i(x)\right) = \frac{1}{N^2}\sum_{i=1}^{N} \mathrm{Var}(Y_i(x)) = \frac{\sigma^2}{N}, \tag{2.4}$$

where $\sigma^2 = \mathrm{Var}\left(Y(x)\right)$, assuming it exists. Equation (2.4) encapsulates the usefulness as well as the limitations of the MC method. First and foremost, it indicates that the order of convergence of the root mean square error (RMSE) is $\mathcal{O}(N^{-1/2})$, which is independent of the dimension d of the random vector Θ. This result contrasts with quadrature approaches whose order of convergence decreases with the dimensions and suffers from the curse of dimensionality. Moreover, this method is very simple to implement, since one only needs to be able to sample from Θ. The downside of the method is that its order of convergence is low ($N^{-1/2}$) and the rate of convergence σ can be high if the second moment of $Y(x)$ is large. In the examples studied in this thesis, we will consider parameter spaces of relatively small dimension, therefore a traditional quadrature would be possible. Nevertheless, in practice, it would be necessary to employ random sampling methods, which is what we have chosen to do.

Latin Hypercube Sampling

Another sampling scheme, the Latin Hypercube Sampling (LHS) consists in stratifying each component of the random vector Θ. Then, the LHS sampling is constructed to ensure that one sample is taken in each subdivision of all components. The LHS scheme necessitates d random permutations π_j of the set $\{0, \ldots, N-1\}$, where d is the dimension of the random vector Θ and N is the number of samples. The collection of N random variables $(T_1, \ldots, T_N)$ are constructed according to

$$(T_i)_j = \frac{\pi_j(i-1) + U_{ij}}{N}, \tag{2.5}$$

where the matrix U features independent coefficients distributed according to $\mathrm{U}(0,1)$. One can show that the random variables T_i are distributed according to $\mathrm{U}(0,1)^d$. Figure 2.1 shows samples obtained with the MC and LHS schemes from a random variable distributed

according to $\mathrm{U}(0,1)^2$. We observe that the MC scheme can create clusters, whereas the LHS scheme ensures that at least one sample is present in each subdivision of the components.

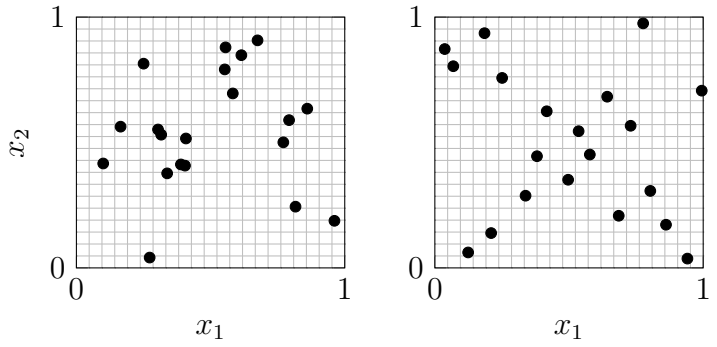


Figure 2.1 Monte Carlo sample (left) and Latin Hypercube sample (right) for $N = 20$.

Because the components of Θ are considered independent, we can easily generate the N samples $\tilde{\Theta}_i$ by the inversion method

$$\begin{aligned}(\tilde{\Theta}_i)_1 =& F^{-1}_{\Theta_1}\left((T_i)_1\right),\\ (\tilde{\Theta}_i)_2 =& F^{-1}_{\Theta_2}\left((T_i)_2\right),\\ &\vdots\\ (\tilde{\Theta}_i)_d =& F^{-1}_{\Theta_d}\left((T_i)_d\right),\end{aligned}$$

where F_{Θ_j} is the cumulative distribution function of the j-th component of Θ. The samples $\tilde{\Theta}_i$ are thus distributed according to Θ. The Latin Hypercube estimator of the expectation $\mathbb{E}(Y(x))$ is

$$\mu_{LHS} = \frac{1}{N}\sum_{i=1}^{N} h(x, \tilde{\Theta}_i). \tag{2.6}$$

The LHS estimator is very similar to the Monte Carlo scheme, apart from the way the samples of Θ are constructed. An interesting result from [49, Proposition 3] states that

$$\mathrm{Var}\left(\mu_{LHS}\right) \leq \frac{\sigma^2}{N-1}, \tag{2.7}$$

where σ^2 is again the variance of $Y(x)$. The variance of the LHS estimator is, in the worst case, equal to that of the MC estimator minus one sample. Hence, the LHS estimator of $\mathbb{E}(Y(x))$ can only perform better than a MC estimator while still being simple to implement. The actual decrease in the variance of the LHS estimator is closely related to the

ANOVA decomposition that will be presented in Section 2.3.

2.3 Sensitivity Analysis

The subject of sensitivity analysis is rich and addresses several types of problems. Roughly speaking, a sensitivity analysis method aims at quantifying the impact of parameters on model outputs. However, the precise objective of the sensitivity analysis differs from one method to another. Also, the quantification of sensitivity often changes with the method adopted. The absence of a unified objective and definition of sensitivity forces the user/modeler to be diligent while performing a sensitivity analysis. The book by Saltelli et al. [50] provides an exhaustive review of the different objectives and methods of sensitivity analysis. Moreover, a thorough discussion about the misconceptions and pitfalls of sensitivity analysis is presented in [51]. The perspective article [52] also provides a useful overview of the objectives and methods associated with sensitivity analysis. To lighten the notation in this section, we will aggregate the input and model parameters in the parameters x, since no particular distinction between the two types of parameters will be considered here. The domain of the functional h is thus $\mathcal{X} \subseteq \mathbb{R}^d$. To ease the presentation, we consider in this section that x_i pertains to the i-th component of x, not another instance of $x \in \mathcal{X}$ as in the remainder of this thesis.

Sensitivity methods are usually classified as either local or global. As their names imply, local methods define a notion of sensitivity at a nominal point $\hat{x} \in \mathcal{X}$. Several such methods are presented in [53, Section 3]. One of the obvious notions of sensitivity consists in computing the partial derivatives $\frac{\partial h}{\partial x_i}(\hat{x})$. From the partial derivative, one can compute the following sensitivity indices

$$D_i = \frac{\frac{\tilde{x}_i}{h(\tilde{x})}\frac{\partial h}{\partial x_i}(\tilde{x})}{\sum_{j=1}^{d}\frac{\tilde{x}_j}{h(\tilde{x})}\frac{\partial h}{\partial x_j}(\tilde{x})}. \tag{2.8}$$

A first normalization makes the notion of sensitivity dimensionless, while a second normalization ensures that $\sum_{i=1}^{d} D_i = 1$. The normalization should be carried out carefully if one of the component is null or if the sum of the indices is null. Based on the sensitivity indices, one can rank the components of x to be the most influential at the nominal point $\tilde{x}$. The sign of D_i also indicates whether the i-th component increases or decreases the functional h at the nominal point $\tilde{x}$. An important and inherent limitation of local sensitivity analysis is that the sensitivity information is only provided at the nominal point $\tilde{x}$. Often, we want a measure of the sensitivity of the functional h over $\mathcal{X}$. A more general setting assimilates the parameters x as random variables/vectors X with distribution μ_X, for which we want to

assess the sensitivity of h over $(\mathcal{X}, \mathcal{B}, \mu_X)$. This is precisely the objective of global sensitivity approaches. For the sake of clarity and conciseness, we will restrict our presentation to two global sensitivity methods, namely the Analysis of Variance (ANOVA) with their Sobol indices and the Active Subspaces method.

ANOVA - Sobol Indices

The Sobol indices are based upon the ANOVA decomposition of the functional h, which allows the decomposition of its variance. More precisely, we seek to attribute the variance of $h(X)$ to each component of X. We refer the interested reader to [54, 55] and [56, Chapter 11] for a detailed description of this decomposition. Before presenting the ANOVA decomposition, we need to introduce some notation. We will note $\mathcal{D} = \{1, 2, \ldots, d\}$, $\mathcal{J} \subseteq \mathcal{D}$, $-\mathcal{J} = \mathcal{D} \setminus \mathcal{J}$, and $X_{\mathcal{J}} = \left(X_{j_1}, X_{j_2}, \ldots, X_{j_{|\mathcal{J}|}}\right)$ for $j_k \in \mathcal{J}$. The underlying hypotheses of the ANOVA decomposition are that X possesses independent components (μ_X is a product measure) and $h(X)$ has finite first and second moments. We can then construct the following 2^d functions

$$
\begin{aligned}
h_0 &= \mathbb{E}(h(X)), && \text{(2.9a)}\\
h_i(x_i) &= \mathbb{E}(h(X)|X_i = x_i) - h_0, && \text{(2.9b)}\\
h_{i,j}(x_i, x_j) &= \mathbb{E}(h(X)|X_i = x_i, X_j = x_j) - h_i(x_i) - h_j(x_j) - h_0, && \text{(2.9c)}\\
&\vdots\\
h_{\mathcal{J}}(x_{\mathcal{J}}) &= \mathbb{E}(h(X)|X_{\mathcal{J}} = x_{\mathcal{J}}) - \sum_{\mathcal{K}\subset\mathcal{J}} h_{\mathcal{K}}(x_{\mathcal{K}}). && \text{(2.9d)}
\end{aligned}
$$

The idea behind these functions is to average the functional h with respect to the parameters $x_{-\mathcal{J}}$ and remove the function that depends on the subset $\mathcal{K} \subset \mathcal{J}$. In this way, the function $h_{\mathcal{J}}$ captures the interaction between the parameters $x_{\mathcal{J}}$. We can show that the expectation of each function is null, apart from h_0. From this property, we can also show that these functions are orthogonal, i.e.,

$$
(h_{\mathcal{K}}, h_{\mathcal{J}}) = \int h_{\mathcal{K}}(x_{\mathcal{K}}) h_{\mathcal{J}}(x_{\mathcal{J}}) \, \mathrm{d}\mu_X = 0 \qquad \text{for } h_{\mathcal{K}} \neq h_{\mathcal{J}}. \tag{2.10}
$$

With the functions (2.9) and their associated properties, we can prove that the functional h can be decomposed as

$$\begin{aligned} h(x) &= h_0 + \sum_{i=1}^{d} h_i(x_i) + \sum_{1\leq i\neq j\leq d} h_{i,j}(x_i, x_j) + \cdots + h_{1,\ldots,d}(x_1, \ldots, x_d) \\ &= \sum_{\mathcal{J}\subseteq\mathcal{D}} h_{\mathcal{J}}(x_{\mathcal{J}}). \end{aligned} \tag{2.11}$$

This is the so-called ANOVA decomposition. The terms h_i (and h_0) are referred to as the additive parts of the functional h. These terms are important because they individually depend on only one parameter. The ANOVA decomposition presents interesting features: first the decomposition involves a finite number of terms; second, each function depends on its own set of parameters. The final and important feature for the purpose of sensitivity analysis is that (2.11) also decomposes the variance

$$\mathrm{Var}\left(h(X)\right) = \sum_{\mathcal{J}\subseteq\mathcal{D}} \mathrm{Var}(h_{\mathcal{J}}(X_{\mathcal{J}})). \tag{2.12}$$

The idea introduced by Sobol [57] is to employ this decomposition of the variance to assess which parameters affect the least or the most the variance. The notion of sensitivity is thus directly related to the variance of $h(X)$. To this end, we derive three types of sensitivity indices

$$S_{\mathcal{J}} = \frac{\mathrm{Var}(h(X_{\mathcal{J}}))}{\mathrm{Var}(h(X))}, \tag{2.13}$$

$$\underline{S}_{\mathcal{J}} = \sum_{\mathcal{K}\subseteq\mathcal{J}} S_{\mathcal{K}}, \tag{2.14}$$

$$\overline{S}_{\mathcal{J}} = \sum_{\mathcal{K}\cap\mathcal{J}\neq\emptyset} S_{\mathcal{K}}. \tag{2.15}$$

The indices $S_{\mathcal{J}}$ indicate the fraction of the variance explained solely by the functions $h_{\mathcal{J}}$. Importantly, the indices do not indicate the total importance of the parameters $x_{\mathcal{J}}$, because the components of $x_{\mathcal{J}}$ also appear in other indices. The indices $\underline{S}_{\mathcal{J}}$ are usually called the closed Sobol indices and aggregate all the variance explained solely by parameters which are subsets of $\mathcal{J}$. The closed indices $\underline{S}_i = S_i$ directly indicate if the parameters X_i are important to the variance of $h(X)$. Indeed, if the index S_i is close to unity, then we can assess that the parameter X_i explains nearly all the variance of $h(X)$. If the sum $\sum_{i=1}^{d} S_i$ is closed to unity, then h is nearly additive, since the variance of h is almost completely explained by the parameters taken individually, without any interaction. However, if the closed index S_i is very small, this does not necessarily indicate that the parameter X_i is not influential with respect

to the variance, since its influence may appear with the interaction of other parameters.

The indices $\overline{S}_{\mathcal{J}}$ are called the total Sobol indices since they encompass all the contributions from the parameters $x_{\mathcal{J}}$, including interactions with the parameters $x_{-\mathcal{J}}$. If a closed index $\overline{S}_i$ is small, then we can assess that the parameter X_i does not contribute in any way to the variance $\mathrm{Var}(h(X))$. The two types of Sobol indices, the closed and total ones, provide different but complementary information about the sensitivity of the model. Again the notion of sensitivity here is the variance $\mathrm{Var}(h(X))$. If the objective is to assess which parameter generates the most uncertainty, then we seek the parameter X_i which possesses a large closed index S_i. This setting is referred to as Factor Prioritization in [50]. On the other hand, if the objective of the sensitivity analysis is to determine which parameters do not drive the variance of $h(X)$, then we seek the total indices $\overline{S}_i$ that are closest to zero.

The computation of the closed and total Sobol indices have been extensively studied in the literature and the interested reader is referred to the article [58] and the references within. The condition that the components of the random variable X are independent is rather important in the derivation of meaningful sensitivity indices. An ANOVA decomposition can be obtained if the random variable X possesses non-independent components [59]. However, the construction of interpretable sensitivity indices from the decomposition is much more difficult. Several approaches have been adopted to address this problem, such as the Generalized Sobol indices [60] or the Shapley value [61].

We conclude the discussion on the Sobol indices and the ANOVA decomposition by revisiting the Latin Hypercube sampling scheme. We can now describe more precisely how the LHS might perform better than a simple MC scheme. Having introduced the ANOVA decomposition (2.11), we can reformulate the result of Corollary 1 in [62] as

$$\mathrm{Var}\left(\mu_{LHS}\right) = \frac{1}{N} \sum_{|\mathcal{J}|>1} \mathrm{Var}\left(h_{\mathcal{J}}(X_{\mathcal{J}})\right) + \mathcal{O}\Big(N^{-1}\Big)\,. \tag{2.16}$$

We observe that the rate of convergence of the LHS sampling depends upon the non-additive terms of the ANOVA decomposition of $h(X)$. Hence, the LHS scheme will perform better if the $h(X)$ is nearly additive, compared to a simple Monte Carlo scheme.

Global Derivative-Based Sensitivity Analysis and Active Subspaces

Global sensitivity analysis methods that employ partial derivative as their notion of sensitivity can also be formulated. For example, the works of Sobol et al. [63] and Lamboni et al. [64]

define the following sensitivity indices

$$\nu_i = \mathbb{E}\left(\left(\frac{\partial h}{\partial x_i}(X)\right)^2\right) = \mathrm{Var}\left(\frac{\partial h}{\partial x_i}(X)\right) + \mathbb{E}\left(\frac{\partial h}{\partial x_i}(X)\right)^2. \tag{2.17}$$

The sensitivity indices ν_i is composed of two terms, the first one indicates how the partial derivative varies according to the measure μ_X and the second term takes into account the magnitude of the mean. The notion of sensitivity that stems from ν_i is different than that of Sobol: we do not seek here to decompose the variance of $h(X)$ and find the components of the parameters X that contribute the most or the less to this variance. The notion of sensitivity captured by ν_i is much more closely related to the shape of the surface of response of h. Even though their notion of sensitivity is different, the indices ν_i can be employed to bound the total Sobol indices, as demonstrated in [63,64]. Naturally, the computation of the sensitivity indices ν_i supposes that $h \in H^1(\mathcal{X}, \mathcal{B}, \mu_X)$, whereas the computation of the Sobol indices requires only that $h \in L^2(\mathcal{X}, \mathcal{B}, \mu_X)$.

Another global sensitivity method that employs derivative information is the Active Subspaces method [65, 66]. The quantity $\frac{\partial h}{\partial x_i}(X)$ introduced previously is a random variable. Thus, $\nabla h(X)$ can be viewed as a random vector, whose components may be non-independent (even if the components of X are considered independent). We can construct the matrix M_h

$$M_h = \mathbb{E}\left(\nabla h(X) \nabla h(X)^T\right) = \mathrm{Cov}\left(\nabla h(X)\right) + \mathbb{E}\left(\nabla h(X)\right) \mathbb{E}\left(\nabla h(X)\right)^T, \tag{2.18}$$

which encodes the covariance matrix of $\nabla h(X)$ as well as its scaling. The central idea of the Active Subspaces method is to perform an eigendecomposition of the positive semi-definite matrix M_h

$$M_h = W \Lambda W^T, \tag{2.19}$$

where W is built from the orthogonal eigenvectors w_i and Λ is the diagonal matrix containing the corresponding positive eigenvalues λ_i (ordered in decreasing order). With the eigendecomposition, we can identify the linear structure of the parameters X that better explains the variation and the scaling of the gradient. More precisely, the eigenvector w_1 indicates the linear combination of the components of X that explained the most the derivative of h, followed by the direction w_2, and so on. Thus, we can employ the Active Subspaces method

to reduce the dimension of the parameters space. For example, we can write

$$\Lambda = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix}, \qquad W = [W_1, W_2], \tag{2.20}$$

where Λ_1 and W_1 contain the m most important eigenvalues and their corresponding eigenvectors. The matrices Λ_2 and W_2 respectively encode the less important eigenvalues and associated eigenvectors. Multiple rules can be chosen to define the value of m, e.g. the value of the first indices k such that $\lambda_k/\lambda_1 \leq \varepsilon$ or the first value of k such that $\sum_{i=1}^{k} \lambda_i / \operatorname{Tr}(M_h) \leq \varepsilon$. We can perform the reduction of the parameter space by defining the function $\tilde{h}$

$$\tilde{h}(x) = \mathbb{E}(h|W_1 W_1^T x) \tag{2.21}$$

which is the conditional expectation of h given the projection of x onto the span of the eigenvectors encoded in W_1. The function $\tilde{h}$ can be viewed as the function that averages out the function h in the directions spanned by the eigenvectors encoded in W_2. Hence, the reduced-order model $\tilde{h}$ approximates the function h while its domain is of dimension $m \leq d$.

Sensitivity indices can also be defined from the eigendecomposition (2.19), as proposed in [67]. The sensitivity indices α_i aggregate the information pertaining to the i-th components via

$$\alpha_i = \sum_{j=1}^{m} \lambda_j (w_j)_i^2. \tag{2.22}$$

The sensitivity indices α_i are bound by the sensitivity indices ν_i and they are equal if $m = d$, meaning that there is no reduction in space. Furthermore, the sensitivity indices α_i can be related to the total Sobol indices [67]. The computation of the sensitivity indices ν_i (2.17) and the matrix M_h (2.18) can be computed by a Monte Carlo or Latin Hypercube sampling, as described in Section 2.2. As a final note, we mention that the Active Subspaces method can be adapted to seek the most important dimensionless quantities [68].

Choice of the Sensitivity Method

We will adopt in Chapter 3 the Active Subspaces notion of sensitivity instead of the variance decomposition as in the Sobol's indices. There are three reasons justifying this choice. First and foremost, the notion of sensitivity of the Active Subspaces reflects, via the use of derivatives, the influence of the parameters on the response surface. The Sobol indices only capture the influence of the parameters on the variance of $h(X)$. We illustrate this assessment with

the following example. Let us consider $X \sim \mathcal{U}(0,1)^2$ and the functional

$$h(x) = \sin(2n\pi x_1) + 10x_2, \tag{2.23}$$

for (hyper-) parameter $n \in \mathbb{N}^+$. Figure 2.2 illustrates the functional h for $n = 1$ and $n = 4$. The influence matrix M_h is

$$M_h = \begin{bmatrix} 2\pi^2 n^2 & 0 \\ 0 & 100 \end{bmatrix}, \tag{2.24}$$

whereas the Sobol indices are $S_1 = \underline{S}_1 = \overline{S}_1 = 3/203$ and $S_2 = \underline{S}_2 = \overline{S}_2 = 200/203$. We observe that the Sobol indices do not depend upon the choice of the (hyper-) parameter n. As Figure 2.2 shows, the frequency n does not affect the overall variance of $h(X)$ and it is mainly the parameter x_2 that explains the variance. On the other hand, the Active Subspaces indices are $\alpha_1 = 2\pi^2 n^2$ and $\alpha_2 = 100$, indicating that the frequency n now affects the importance of the first component x_1. Indeed, the partial derivative $\partial h/\partial x_1$ increases in magnitude as n increases and this is captured by α_1. By capturing the covariance and expected magnitude of the gradients, the Active Subspaces method indirectly captures how the components affect the response surface.

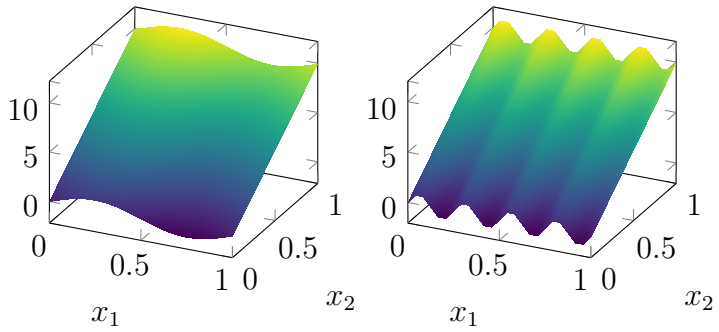


Figure 2.2 Function (2.23) for $n = 1$ (left) and for $n = 4$ (right).

The second reason justifying the use of the Active Subspaces is that we want the global sensitivity method to reduce to a local sensitivity method if the uncertainty on X vanishes. More precisely, in the case where $X \sim \delta_{x_0}$, with δ_{x_0} the delta Dirac distribution centered at x_0, we still want to assess the sensitivity of $h(X)$. The Sobol indices cannot be employed in this case, because there is no longer a variance to decompose. On the other hand, the Active Subspaces method reduces to a local sensitivity method, where the absolute value of the gradient is taken as a notion of sensitivity.

The third and final reason concerns the possibility that the components of X may be non-

independent. This case greatly complicates the use and interpretation of the Sobol indices, as mentioned previously. On the other hand, the Active Subspaces method can handle the non-independence of the components of X without any modification or complication (apart from the necessity to sample X).

2.4 Gaussian Processes

Before proceeding with the presentation of the calibration and validation problems, we need to provide a brief overview of stochastic processes. Sometimes, a collection of random variables $\{Y(x)\}_{x\in\mathcal{X}}$ will possess a dependence structure and such object will be assimilated as a stochastic process. For the sake of clarity in the presentation, we will refer to a stochastic process $\{Y(x)\}_{x\in\mathcal{X}}$ simply as Y. It would be clear from the context whether Y refers to a random variable or a stochastic process. Moreover, for the purpose of the thesis, we will restrict our presentation of stochastic processes to Gaussian processes. They will be employed in Chapter 4 to represent modeling discrepancy and the corrected model predictions. The following presentation is essentially inspired by [69]. The Gaussian process Y is described as

$$Y \sim \mathcal{GP}(m(x), k(x, x')), \tag{2.25}$$

where m and k are the mean and kernel function of the Gaussian process. An important feature of Gaussian processes is that they are fully described by their mean and kernel functions. The mean function m encodes how the trend of the stochastic process varies in the space $\mathcal{X}$. On the other hand, the kernel function k specifies the behavior and regularity of the function $Y(\cdot, \omega)$. If the mean function does not depend upon x and the kernel function can be written as a function of $|x - x'|$, then the process is considered stationary, since its behavior does not depend upon the point $x \in \mathcal{X}$. Otherwise, the Gaussian process is non-stationary. This particular distinction will be relevant in the thesis. Let us consider the three Gaussian processes $Y_i \sim \mathcal{GP}(m_i(x), k_i(x, x'))$, for $i = 1, \ldots, 3$, with

$$\begin{aligned}
&m_1(x) = 0, && k_1(x, x') = \sigma^2 \exp\left(-\frac{r^2}{2\ell^2}\right),\\
&m_2(x) = 1 - x, && k_2(x, x') = \sigma^2 \exp\left(-\frac{r^2}{2\ell^2}\right),\\
&m_3(x) = 0, && k_4(x, x') = \sigma^2 \left(1 + \frac{\sqrt{3}r}{\ell}\right) \exp\left(-\frac{\sqrt{3}r}{\ell}\right),
\end{aligned}$$

where $r = |x - x'|$. The term σ^2 defines the variance of the Gaussian processes and is set to $\sigma^2 = 1$. The term ℓ controls the length-scale of the functions. The first kernel will be

employed extensively in Chapter 4. The Gaussian processes are displayed in Figure 2.3 as well as some of their realizations. The first Gaussian process has a mean zero trend and the choice of a radial basis function kernel allows the representation of smooth functions of length-scale ℓ. The second Gaussian process possesses the same kernel as the first one, but it exhibits a polynomial trend, making it non-stationary. The last Gaussian process is stationary with a Matérn kernel ($\nu = 3/2$) which is mean-square continuous and differentiable. The choice of $\nu = 3/2$ allows the representation of functions that possess some regularity while not being smooth.

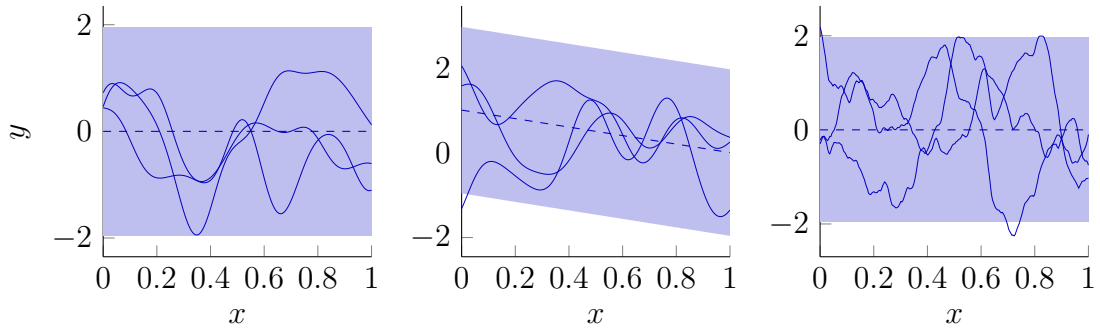


Figure 2.3 Illustration of Gaussian processes. The dashed blue lines indicate the mean function whereas the shaded areas represent the 95% prediction interval. Three realizations of the Gaussian processes are displayed by thin lines.

We can extract a multidimensional random normal variable $\hat{Y}$ by evaluating the Gaussian process $Y \sim \mathcal{GP}(m(x), k(x, x'))$ at a finite set of points $\{x_1, \ldots, x_N\}$. The normal variable $\hat{Y}$ is described by the following first and second moments

$$\mathbb{E}(\hat{Y}) = (m(x_1), \ldots, m(x_N)), \tag{2.26}$$

$$(\mathrm{Cov}(\hat{Y}))_{i,j} = k(x_i, x_j). \tag{2.27}$$

We will construct several such random normal variables in Chapter 4 to compute (log)likelihoods. The calculation of the (log)likelihoods involves the inverse of $\mathrm{Cov}(\hat{Y})$, which is computed with a Cholesky decomposition. It is important to note that depending upon the choice of the kernel function, the covariance matrix $\mathrm{Cov}\,\hat{Y}$ may be full. Hence, if the number of data considered is very large, the solution of linear systems involving the covariance matrix must be dealt with other approaches, such as those described in [69, Chapter 8]. However, for the examples considered in the thesis, the data set are relatively small, so a Cholesky decomposition is sufficient.

A useful way to create a non-stationary Gaussian process from a stationary one is mentioned

in [69, Section 4.2.3] and further detailed in [70,71]. The process can be made non-stationary by computing the distance r using the Mahalanobis distance

$$q(x, x') = (x - x')^T \left(\frac{\Sigma(x) + \Sigma(x')}{2} \right)^{-1} (x - x'), \tag{2.28}$$

for some function $\Sigma : \mathcal{X} \to \mathcal{S}^d_{++}$, where $\mathcal{S}^d_{++}$ is the set of positive definite matrices of size d. Moreover, one can directly alter the variance of the Gaussian process by introducing a strictly positive function $\sigma : \mathcal{X} \to \mathbb{R}^+$. As proposed in [70,71], one can use radial basis functions to create the non-stationary Gaussian process with kernel function $\tilde{k}$

$$\tilde{k}(x, x') = \sigma(x)\sigma(x')|\Sigma(x)|^{1/4}|\Sigma(x')|^{1/4} \left| \frac{\Sigma(x) + \Sigma(x')}{2} \right|^{-1/2} \exp(-q(x, x')), \tag{2.29}$$

where the normalization ensures that the standard deviation is provided directly by the function σ. As an illustrative example, let us consider two mean zero non-stationary Gaussian processes obtained with

$$\begin{aligned} \sigma_1(x) &= 0.1 + \sin(2\pi x)^2 & \Sigma_1(x) &= 1, \\ \sigma_2(x) &= 1 & \Sigma_2(x) &= \exp(-10 + 20x). \end{aligned}$$

Figure 2.4 illustrates these two non-stationary Gaussian processes. For the first one, we clearly observe that the variance of the Gaussian process varies with x whereas its length-scale remains constant. On the other hand, the second non-stationary Gaussian process possesses a constant variance, but its length-scale increases with x.

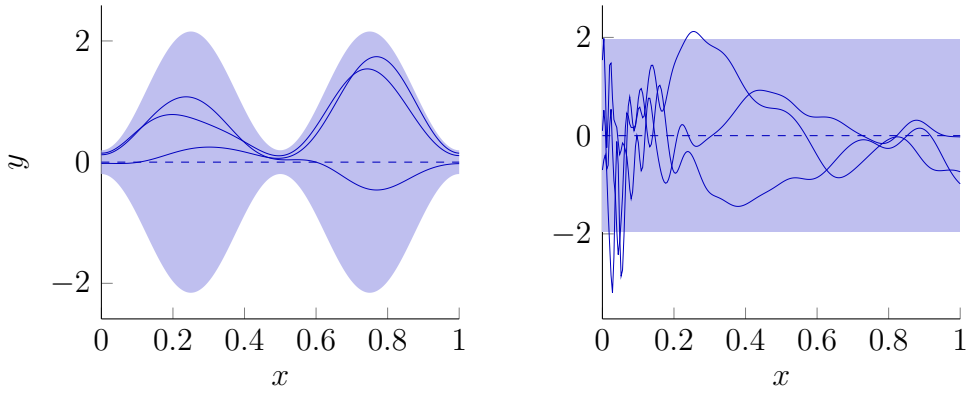


Figure 2.4 Illustration of non-stationary Gaussian processes constructed from the Gaussian kernel. The dashed blue lines indicate the mean function whereas the shaded areas represent the 95% prediction interval. Three realizations of the Gaussian processes are displayed by thin lines.

2.5 Brief Overview of Optimization Methods

Several problems in the thesis will be cast as optimization problems for which an analytical solution is not known. We then resort to optimization methods to solve these problems. It is not our objective here to provide a comprehensive review and taxonomy of optimization problems and optimization algorithms; such task is far beyond the scope of the thesis. Rather, we will describe more precisely two types of optimization problems encountered in the thesis. The features and information we possess regarding these two optimization problems will guide our choice of optimization methods employed to solve them. Without loss of generality, a constrained optimization problem can be written as

$$\min_{x \in \mathbb{R}^d} \quad h(x), \tag{2.30a}$$

$$\text{s.t.} \quad c_{\mathcal{E}}(x) = 0, \tag{2.30b}$$

$$c_{\mathcal{I}}(x) \geq 0, \tag{2.30c}$$

where h is the objective functional, $c_{\mathcal{E}}$ encodes the m_e equality constraints, and $c_{\mathcal{I}}$ encodes the m_i inequality constraints. To lighten the notation, the inequalities are to be thought as component-wise. We will note the feasible set $\tilde{\mathcal{X}} = \{x \in \mathbb{R}^d \,|\, c_{\mathcal{E}}(x) = 0\,,\, c_{\mathcal{I}}(x) \leq 0\}$ and the global solution to (2.30) as $x^{**} \in \tilde{\mathcal{X}}$. More often than not, we will obtain a local solution of (2.30) denoted by x^*. The computation of a global solution is a very challenging task, since it supposes some knowledge of the value of h over $\tilde{\mathcal{X}}$, or subset of it. We will employ here two local optimization methods that compute a local solution x^* to (2.30) and will briefly discuss how these methods attempt to find a global solution x^{**}.

A first type of optimization problems, employed for the calibration of the modeling discrepancy at Chapter 4, is the optimization of a continuous differentiable function with differentiable constraints. The presence of constraints reduces the scope of the optimization algorithms available, since the solution must be found within the feasible region $\tilde{\mathcal{X}} \subseteq \mathbb{R}^d$. The fact that we can access the first-order (and possibly second-order) information about the functions h, $c_{\mathcal{E}}$, and $c_{\mathcal{I}}$, provide crucial information that can be exploited to speed up the convergence to a local solution x^*. To the best of the author's knowledge, there exist two types of optimization algorithms that can handle the constraints and exploit derivative information, namely the active-set method and the interior-point method. These types of algorithms can handle optimization problems with a lot of parameters, but the problem studied in the thesis will remain of small size. We will employ here the Knitro [72] and IPOPT [73] solvers that implement versions of an interior-point method.

A second type of optimization problems, arising in both Chapter 3 and Chapter 4, is commonly referred to as a black-box constrained optimization problem. Compared to the first type of optimization problems, we do not have prior information about the objective function h and the constraint functions $c_{\mathcal{E}}$ and $c_{\mathcal{I}}$. We can only evaluate the value of $h(x)$, $c_{\mathcal{E}}(x)$, and $c_{\mathcal{I}}(x)$ by interrogating the black-box with the input x. Importantly, we do not have information about the derivatives of these functions. The black-box solver that will be employed to solve such optimization problems is the NOMAD solver [74] that implements the Mesh Adaptive Direct Search (MADS) algorithm [75]. In contrast with the interior-point methods, MADS can handle small optimization problems, which is the case of the problem considered in the thesis.

Interior-Point Method

We provide here a very general presentation of the interior-point method to solve differentiable non-linear constrained optimization problems. The goal is to highlight the main idea without delving into the fine details. Indeed, IPOPT implements its own version of the interior-point method, while Knitro provides two different implementations. The following presentation is largely inspired by the book of Nocedal and Wright [76].

The fact that the first type of optimization problems are differentiable allows us to employ the Karush-Kuhn-Tucker (KKT) conditions [76, Theorem 12.1], which prescribe the necessary conditions for a local minimum. We first need to introduce the Lagrangian

$$L(x, \lambda_e, \lambda_i) = h(x) - \lambda_e^T c_{\mathcal{E}}(x) - \lambda_i^T c_{\mathcal{I}}(x), \tag{2.31}$$

where $\lambda_e \in \mathbb{R}^{m_e}$ and $\lambda_i \in \mathbb{R}^{m_i}$ are the Lagrange multipliers for the equality and inequality constraints, respectively. With the Lagrangian, we can write the KKT conditions as

$$\nabla_x L(x^*, \lambda_e^*, \lambda_i^*) = 0, \tag{2.32a}$$
$$c_{\mathcal{E}}(x^*) = 0, \tag{2.32b}$$
$$c_{\mathcal{I}}(x^*) \geq 0, \tag{2.32c}$$
$$\lambda_i^* \geq 0, \tag{2.32d}$$
$$\Lambda_i^* c_{\mathcal{I}}(x^*) = 0, \tag{2.32e}$$

where Λ_i is a diagonal matrix whose entries are λ_i^*. Equation (2.32e) is called the complementary condition, which indicates whether or not the inequality constraints are active or not. If a component of λ_i is not zero, it indicates that the corresponding inequality constraint is

active. The challenge of employing the KKT conditions (2.32) to construct optimization algorithms lies in the combinatorial nature of Equation (2.32e). We do not know a priori which inequality constraints are active or inactive, and exploring the 2^{m_i} possibilities is unfeasible.

The core idea of the interior-point method is to incorporate the inequality constraints into the objective function via a log-barrier. The optimization problem (2.30) is then recast as

$$\min_{x\in\mathbb{R}^d} \quad h(x) - \mu \sum_{i=1}^{m_i} \log(s_i), \tag{2.33a}$$

$$\text{s.t.} \quad c_{\mathcal{E}}(x) = 0, \tag{2.33b}$$

$$c_{\mathcal{I}}(x) - s = 0, \tag{2.33c}$$

where μ is a positive term that should be driven to zero and s represents the slack variables. This way, the combinatorial problem is circumvented. The KKT conditions of Problem (2.33) then read

$$\nabla h(x) - J_{c_{\mathcal{E}}}^T \lambda_e - J_{c_{\mathcal{I}}}^T \tilde{\lambda}_i = 0, \tag{2.34a}$$

$$S\tilde{\lambda}_i = \mu e, \tag{2.34b}$$

$$c_{\mathcal{E}}(x) = 0, \tag{2.34c}$$

$$c_{\mathcal{I}}(x) - s = 0, \tag{2.34d}$$

where $J_{c_{\mathcal{E}}}$ and $J_{c_{\mathcal{I}}}$ are the Jacobian of the equality and inequality constraints, respectively. The matrix S is a diagonal matrix whose entries are the slack variables s and e is a vector of ones. We can observe from the KKT conditions (2.34) that the complementary conditions have been relaxed by the term μe. An important aspect of an interior-point algorithm resides in the strategy adopted to decrease the parameter μ. It can be decreased monotonically each time a solution to (2.34) is found (a solution that is a minimum of (2.33)) or it can be modified adaptively [76, Section 19.3]. With a strategy to decrease μ, we can adopt two approaches to solve (2.33). The first approach is the line-search paradigm, where a direction of descent is computed by solving the KKT conditions (2.34) in a Newton fashion. Then a step is computed to ensure that some functional of the objective function, constraints violation, and complementary value, is decreased. The second paradigm is the trust-region, where a sequence of quadratic versions of (2.34) are solved. The interested reader is referred to [76, Chapter 19] for a full and comprehensive description of each variant of the interior-point method.

Any interior-point algorithm provides a local solution x^* to the optimization problem (2.30). However, the optimization problems studied in the thesis are not convex, hence a local

solution is not guaranteed to be a global solution x^{**}. By employing a local optimization solver for a set of different initial points, we can obtain a set of local solutions that hopefully contains a global solution. This procedure is readily implemented in the solver Knitro, which defines the initial guesses in a Monte Carlo fashion. There is no guarantee that a global solution is computed by using this strategy, but at least different regions of the parameter space are explored. This procedure will be utilized in the thesis since we have observed that the first type of optimization problems possess multiple local minima.

Mesh Adaptive Direct Search Algorithm

As mentioned earlier, the second type of optimization problems considered here can be collectively coined as black-box optimization problems. Because the information on derivatives is not accessible, we cannot exploit the first-order conditions (2.32) to verify the stationarity of a point, nor construct optimization algorithms. An alternative idea consists in constructing a mesh around the iterates and explore candidates on it. In order to ensure that a local solution is obtained, the mesh and its exploration must be carefully designed. The MADS algorithm ensures that a local minimum can be found. The following discussion is a very short summary of the work presented in [75, 77] and is largely inspired by the presentation made in [78].

In a nutshell, the MADS algorithm is separated in two sequential steps, the first one is the search step and the second one is the poll step. We will present the search step at the end, because it pertains to a global search and is not crucial for convergence to a local minimum. To better describe the poll step, let us consider an iterate $x^{(k)}$. The poll step constructs a mesh and a frame around the current iterate $x^{(k)}$, as shown in Figure 2.5. An important aspect of the algorithm is that the frame is greater than the size of the mesh.

Once the frame and the mesh are constructed, a set of candidate points $\mathcal{P}_K$ are evaluated. The choice and number of candidates in the set $\mathcal{P}_k$ are not arbitrary. More precisely, the set is constructed according to

$$\mathcal{P}_k = \{x_k + d \,|\, d \in \mathcal{D}_k\}, \tag{2.35}$$

where $\mathcal{D}_k$ is a positive spanning set of $\mathbb{R}^d$. A positive spanning set of $\mathbb{R}^d$ consists of a finite collection of elements of $\mathbb{R}^d$ whose positive linear combinations span $\mathbb{R}^d$. A last restriction on the set $\mathcal{D}_k$ is that the directions must arrive at vertices of the mesh inside the frame. The objective function and the constraints are evaluated at the candidate points of $\mathcal{P}_k$. If a better candidate is found, then it is chosen as the next iterate. Otherwise, the mesh and frame are

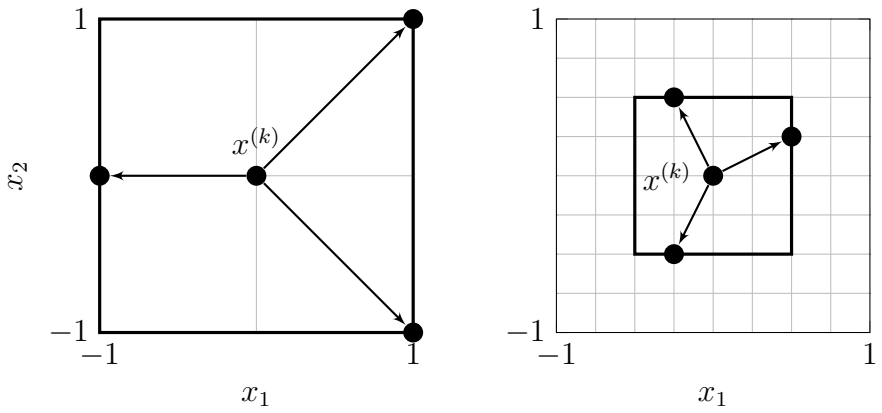


Figure 2.5 Illustration of two poll steps of the MADS algorithm. The two frames are depicted by thick squares and the candidate points are shown as points. If no candidate points decrease the objective function (left), the frame and mesh are reduced and new candidate points are considered (right).

both reduced, but the frame by a greater factor. Then, the poll step is repeated again, with a new positive spanning set of directions, as shown in Figure 2.5. With this algorithm, the set of normalized directions explored around the best candidates grow denser as the mesh and frame sizes are decreased. Depending upon the regularity of h and c, it can be shown that the asymptotic solution is a local minimum [75].

The MADS algorithm handles inequality constraints via an extreme or progressive barrier. The extreme barrier simply rejects the points that do not satisfy the condition $c_{\mathcal{I}} \geq 0$. The progressive barrier p introduces the function

$$p(x) = \begin{cases} \sum_{i=1}^{m_i} (\max\{-c_{\mathcal{I},i}, 0\})^2 & \text{if } x \in \hat{\mathcal{X}} \\ \infty & \text{otherwise} \end{cases} \tag{2.36}$$

where $\hat{\mathcal{X}}$ defines the set of parameters x for which the evaluation of the functions h, $c_{\mathcal{I}}$, and $c_{\mathcal{E}}$ are possible, i.e., the black-box does not return an error. With the progressive barrier, the MADS algorithm keeps track of feasible and infeasible iterates (a feasible solution x is such that $p(x) = 0$). The poll step is applied to both feasible and infeasible iterates, with specific rules to update these points according to the value of the objective function h and the progressive barrier p. To the best of the author's knowledge, the MADS algorithm can handle linear equality constraints, essentially by exploring only directions that are in the nullspace.

The search step of the MADS algorithm allows the exploration of the parameter space that is not located on the frame. The search step can thus be viewed as a way to seek better

local solutions in the hope to find a global minimum. Multiple strategies can be employed to perform this global exploration, such as the Variable Neighborhood Search [79]. The search strategy can even be heuristic, since it is the poll step that ensures convergence to a local minimum. If a better candidate is found during the search phase, the current iterate is replaced by this better candidate.

2.6 Inference/Calibration Methods

The present section and Sections 2.7 and 2.8 bridge the abstract model with the reality via experimental data. Throughout the thesis, we view experimental data as the only way to gather information about a system of interest. We note that the process of gathering experimental data usually relies on some models, i.e., the relation between a sensor input and output. However, we will consider that these metrology models are perfect, or that their modeling errors are negligible with respect to that of the model describing the whole system of interest.

We now provide a description of the experimental design. As mentioned in Section 2.1, the parameters x simultaneously act as input of the model and of the system of interest. More precisely, the parameters x describe the conditions at which the system of interest is studied as well as the spatio-temporal locations of the sensors. In Chapter 3, we will need to further refine the definition of the inputs x, leading to some notation discrepancies.

Each experiment is described by an experimental design $x_{\text{exp}} = (x_1, \dots, x_m) \in \mathbb{R}^{d \times m}$, where d is the dimension of $\mathcal{X}$ and m is the number of experimental data recorded. Associated with the experimental design is the collection of experimental data $y_{\text{exp}} = (y_1, \dots, y_m) \in \mathbb{R}^m$. For the sake of the discussion, we will suppose that there is only one type of experimental data recorded at the experimental design. When appropriate, we will include a discussion regarding the possibility of obtaining experimental data of different nature. We will often consider a sequence of N experiments that will be described as $(x_{\text{exp},1}, y_{\text{exp},1}), \dots, (x_{\text{exp},N}, y_{\text{exp},N})$.

We assimilate the experimental data y_{exp} as a realization of the random variable Y_{exp}. We emphasize that the precise description of Y_{exp} is elusive because it pertains to the system of interest. It is the whole objective of the statistical modeling to approximate the random process Y_{exp}. For example, one prevalent way to statistically model the system of interest is

$$Y(x) = h(x, \theta) + \Phi(x), \tag{2.37}$$

where Y is the stochastic process indexed by the parameters x, $h(x, \theta)$ are the model predictions at x, and Φ is the observation noise, a stochastic process also indexed by x. Usually,

the observation noise is supposed to be a white-noise, whose variance is that of the sensor recording the experimental data. With (2.37) at hand, we can define more precisely the calibration process. The calibration process consists in determining the model parameters θ such that the realizations y_{exp} can reasonably arise from Y. In the highly idealized case of a perfect model h, we want to compute θ such that $Y_{\text{exp}} = (Y(x_1), \dots, Y(x_m))$. At first, we will consider this idealized setting to present calibration approaches. We will revisit this assumption in Section 2.7.

Maximum Likelihood Estimation

The first calibration paradigm that we present here is the Maximum Likelihood Estimation (MLE). The following discussion is inspired by the presentation made in [12, 13]. It is based upon the likelihood function

$$L(\theta) = f_{Y|\theta,x_{\text{exp}}}(y_{\text{exp}}), \tag{2.38}$$

where $f_{Y|\theta,x_{\text{exp}}}$ is the conditional density of Y given the model parameters θ and the experimental design x_{exp}. A sequence of N independent experiments consists in obtaining the realizations $y_{\text{exp},1}, \dots, y_{\text{exp},N}$ of the independently random variables $Y_1, \dots, Y_N$ that may or may not be identically distributed. With a sequence of independent experiments, the likelihood functions reads

$$L(\theta) = \prod_{i=1}^{N} f_{Y_i|\theta,x_{\text{exp},i}}(y_{\text{exp},i}). \tag{2.39}$$

The maximum likelihood estimator θ^* is defined as

$$\theta^* \in \underset{\theta\in\mathcal{T}}{\text{argmax}}\, L(\theta). \tag{2.40}$$

The MLE estimator possesses very useful features that we detail here. First, the maximization problem (2.40) provides a methodology to derive an estimator of θ that does not depend upon the model h considered. Hence, we do not need to construct our own estimator of θ. For example, in Chapter 4, we will extensively employ the MLE Problem (2.40) under variants of the representation (2.37). For the sake of simplicity in the presentation, we present the MLE problem under the assumption that the observation process $\Phi(x) \sim \mathcal{GP}(m(x), k(x, x'))$

$$\theta^* \in \underset{\theta\in\mathcal{T}}{\text{argmax}}\, \frac{1}{2} \sum_{i=1}^{N} \|y_{\text{exp},i} - h(x_{\text{exp},i}, \theta) + m(x_{\text{exp},i})\|^2_{\Sigma^{-1}_{\Phi,i}}, \tag{2.41}$$

where $h(x_{\text{exp},i}, \theta)$ and $m(x_{\text{exp},i})$ are m_i dimensional vectors whose components are computed with each element of the experimental design $x_{\text{exp},i}$. The $m_i \times m_i$ covariance matrix $\Sigma_{\Phi,i}$ is obtained in the same way as in (2.27). For the sake of clarity and conciseness, we employ the notation $\|v\|_B^2 = (B^{1/2}v, B^{1/2}v) = v^T B v$ for $v \in \mathbb{R}^d$ and B a positive definite matrix of size d. Thus, the MLE Problem (2.41) consists in a Least Squares problem, whose weights are given by the covariance matrices $\Sigma_{\Phi,i}$.

The distribution of the MLE estimator is in general not known. Only in the simplest settings, such as when h is linear with respect to the model parameters θ, it is possible to know the distribution of the MLE estimator. Nevertheless, it is important to observe that the asymptotic properties of the MLE estimator are always well-known, even though they might be difficult to compute for small sample sizes. If we were to design our own estimator, we would need to study these asymptotic properties, which can prove to be far from trivial. Before describing the asymptotic properties of the MLE estimator, we need to introduce some hypotheses. We do not pretend here to provide a minimal nor a full list of the hypotheses since such a task will obfuscate the discussion. Rather, we want to highlight the important conditions relevant to our presentation. The following list is an aggregation of those presented in [12, Definition 2.78], [13, Theorem 16.2], and [80, Chapter 7].

1. The statistical model is perfect, i.e., there exists a value of the model parameters θ such that Y_{exp} can be exactly described by the statistical model;
2. The model parameters are identifiable, i.e., $f_{Y|\theta,x_{\text{exp}}} = f_{Y|\theta',x_{\text{exp}}} \Leftrightarrow \theta = \theta'$;
3. The set $\mathcal{S} = \{y \mid f_{Y|\theta,x_{\text{exp}}}(y) > 0\}$ does not depend upon θ;
4. $\forall y \in \mathcal{S}$, $\frac{\partial}{\partial\theta} f_{Y|\theta,x_{\text{exp}}}(y)$, $\frac{\partial^2}{\partial\theta^2} f_{Y|\theta,x_{\text{exp}}}(y)$, and $\frac{\partial^3}{\partial\theta^3} f_{Y|\theta,x_{\text{exp}}}(y)$ exist;
5. The integral $\int f_{Y|\theta,x_{\text{exp}}}(y)\, \mathrm{d}\mu_Y(y)$ can be differentiated with respect to the model parameters θ.

The first condition is almost always implicitly stated in the literature. It is often coined as: the distribution of Y_{exp} belongs to the parametric family whose density can be described by $f_{Y|\theta,x_{\text{exp}}}$. The second condition is the identifiability condition ensuring that the MLE is unique. This condition can be difficult to assess for complex physical models h, as it may also depends upon the experimental design x_{exp}. The interested reader is referred to [81] for an extensive discussion of the identifiability issue. However, we will revisit this issue in Chapter 4. The third condition is not necessarily restrictive in our context since the model parameters θ affect the deterministic model, not the observation noise. The fourth and fifth

conditions stipulate that the model h must possess some regularity with respect to the model parameters θ, which is not really a restriction since realistically it is the opposite that would require justification on the part of the modeler.

Under these conditions, the asymptotic consistency of the MLE estimator can be proven, see [13, Theorem 16.2] and [80, 7.1.1]. More precisely, let $Y_1, Y_2, \ldots, Y_N$ be a sequence of independent and identically distributed (iid) random variables according to the distribution whose density is $f_{Y|\theta_0,x_{\text{exp}}}$. The model parameters θ_0 denote the true model parameters. Then, the MLE estimator θ_N^* (2.40) is asymptotically consistent, i.e., $\forall \varepsilon > 0$

$$\lim_{N\to\infty} \Pr\left(|\theta_N^* - \theta_0| > \varepsilon\right) = 0. \tag{2.42}$$

The convergence in probability is useful since it indicates that as the number of experiments increase, the estimator should approach the true value θ_0. Again, this result is asymptotic and does not provide information about the quality of the estimator θ^* for finite experimental data.

Another important property of the MLE estimator is that, under the conditions mentioned earlier, it is asymptotically normal. The result can be found in [12, Theorem 7.63], [13, Theorem 16.3], [80, Theorem 7.3.1]. It states that the distribution of the random variable $\sqrt{N}(\theta_N^* - \theta_0)$ converges to the distribution $\mathcal{N}(0, \mathcal{I}^{-1}(\theta_0))$. The matrix I is the Fisher information matrix, which is defined as

$$\mathcal{I}(\theta) = \text{Cov}\left(\nabla_\theta \ln f_{Y|\theta,x_{\text{exp}}}(Y)\right) = -\,\mathbb{E}\left(H_\theta \ln f_{Y|\theta_0,x_{\text{exp}}}(Y)\right), \tag{2.43}$$

where H_θ denotes the Hessian matrix with respect to θ. In Section 3.2.6, we will explicit the dependence of the Fisher Information matrix upon the experimental design x_{exp}. The Fisher information matrix shares some similarities with the influence matrix (2.18). It measures the sensitivity of the log-likelihood with respect to the model parameters. It is interesting to observe that if a component of θ greatly influences the log-likelihood, then the associated variance of the estimator will be small. This normality result holds asymptotically, but it could be exploited to estimate a confidence region for the MLE estimator.

The asymptotic consistency and normality constitute the principal reasons the MLE is widely used. Indeed, we now have the confidence that, as the number of data increases, the MLE estimator will converge to the true underlying model parameters θ_0.

Bayesian Inference

Bayesian inference provides an alternative paradigm to perform calibration, where the model parameters are now considered random and noted Θ. Prior knowledge and assumptions about the model parameters Θ are encoded in the density f_Θ. The prior density f_Θ is updated by the gathering of experimental data $Y_{\text{exp}} = y_{\text{exp}}$ via the likelihood function L. The posterior density $f_{\Theta|Y_{\text{exp}}=y_{\text{exp}},x_{\text{exp}}}$ is computed according to

$$f_{\Theta|Y=y_{\text{exp}},x_{\text{exp}}}(\theta) = \frac{f_{Y|\Theta=\theta,x_{\text{exp}}}(y_{\text{exp}})f_\Theta(\theta)}{\int f_{Y|\Theta=\theta',x_{\text{exp}}}(y_{\text{exp}})f_\Theta(\theta')\,\mathrm{d}\theta'}. \tag{2.44}$$

The posterior density yields the probability of the model parameters $\Theta = \theta$ given the experimental data $Y = y_{\text{exp}}$ and the experimental design x_{exp}. The denominator in (2.44) refers to the model evidence

$$f_{Y|x_{\text{exp}}}(y_{\text{exp}}) = \int f_{Y|\Theta=\theta',x_{\text{exp}}}(y_{\text{exp}})f_\Theta(\theta')\,\mathrm{d}\theta', \tag{2.45}$$

i.e., the density of the experimental data given the experimental design x_{exp} and the model. This model evidence is also referred to as the marginal likelihood. The computation of the model evidence presents several challenges, as highlighted by Llorente et al. [34].

An estimator of the model parameters arising from the Bayesian approach is the Maximum a posteriori (MAP)

$$\theta^*_{\text{MAP}} \in \operatorname*{argmax}_{\theta\in\mathcal{T}} f_{\Theta|Y=y_{\text{exp}},x_{\text{exp}}}(\theta). \tag{2.46}$$

The MAP problem can be viewed as a regularized version of the MLE problem (2.40). Indeed, the MAP problem can be cast as

$$\theta^*_{\text{MAP}} \in \operatorname*{argmax}_{\theta\in\mathcal{T}} \log L(\theta) + \log f_\Theta(\theta). \tag{2.47}$$

The well-posedness of the computation of the posterior (2.44) has been studied in [16, 17], even in the infinite dimensional setting. By well-posedness, we mean that a unique posterior exists, which is moreover stable with respect to perturbations on the experimental data y_{exp}. These results are interesting since they are not asymptotic. However, they do not indicate if the posterior provides information about the true value θ_0. The Bernstein-von-Mises Theorem actually gives asymptotic results for the Bayesian inference [12, Section 7.4], [13, Chapter 20], and [56, Chapter 10]. In addition to the conditions enunciated for the MLE, the prior density f_Θ now needs to be smooth and strictly positive around the true value θ_0. Under

these assumptions, it can be shown that the estimator (2.46) is asymptotically consistent. Moreover, asymptotic normality can be shown, i.e., the conditional distribution of the random variable $\sqrt{n}(\Theta - \theta_0)$ tends to the distribution of a random variable $\mathcal{N}(0, \mathcal{I}^{-1}(\theta_0))$. The precise description of these results can be found in the aforementioned references. The key conclusion from these asymptotic properties is that the influence of the prior f_Θ vanishes as the number of experimental data increases.

In the thesis, we will consider the MLE framework for the calibration task. We justify this choice by the fact that we view the MAP as a regularized version of the MLE. The regularization is provided by the prior knowledge on the model parameters θ. As discussed in further detail in Chapter 4, we wish to calibrate the modeling discrepancy parameters. Since the modeling errors are a priori not known, we do not have prior knowledge on the modeling discrepancy parameters. Hence, the use of the Bayesian framework does not appear beneficial here, because of our inability to provide an informative prior. If, on the other hand, an informative prior on the modeling discrepancy parameters were available, then the analysis in this research could be undertaken just as easily within the Bayesian framework.

2.7 Modeling Errors

We now deviate from the idealized case in which the model can perfectly capture the response of the system of interest. Hence, the discrepancies between the model predictions and the observations cannot solely be explained by the observation errors. The remaining discrepancies will be assimilated as modeling discrepancy. Taking into account the inevitable modeling errors in both the observables and the QoIs constitutes a backdrop of the thesis. It will influence how we approach the calibration and validation tasks.

Calibration with Modeling Errors

The presence of modeling errors implies that the statistical representation Y of the reality Y_{exp} is now misspecified. In this thesis, we will suppose that the misspecification arises from misspecifications of the model h, not from misspecifications of the observation noise. Then, the first condition enunciated in Section 2.6 no longer holds. In turn, this affects the property of the MLE estimator, since the notion of true model parameters θ_0 makes little sense here. The following discussion and results are summarized from [14].

Since the model is not a perfect description of the reality, we now need to differentiate the true process Y_{exp} from our statistical model Y (deterministic model h plus the observation noise Φ). We will denote the distribution and density of the reality Y_{exp} by $\mu_{Y_{\text{exp}}}$ and $f_{Y_{\text{exp}}}$,

respectively. In the same way, we will note the distribution and density of Y respectively as $\mu_{Y|\theta}$ and $f_{Y|\theta}$. We first introduce the Kullback-Leibler divergence $\mathrm{KL}(Y_{\mathrm{exp}}|Y;\theta)$

$$\mathrm{KL}(Y_{\mathrm{exp}}, Y;\theta) = \int \ln\left(\frac{f_{Y_{\mathrm{exp}}}(y)}{f_{Y|\theta}(y)}\right) f_{Y_{\mathrm{exp}}}(y)\,\mathrm{d}y, \tag{2.48}$$

which, roughly speaking, specifies the amount of information needed to describe a sample of Y_{exp} with the statistical model Y. If the model is perfect, then $\mathrm{KL}(Y_{\mathrm{exp}}, Y;\theta_0) = 0$, because the statistical model can describe an experiment on the system of interest. Otherwise, we denote by $\hat{\theta}$ the value of the model parameters that minimizes (2.48), i.e.,

$$\hat{\theta} \in \underset{\theta\in\mathcal{T}}{\mathrm{argmin}}\, \mathrm{KL}(Y_{\mathrm{exp}}, Y;\theta). \tag{2.49}$$

The MLE estimator (2.40) can be shown to be an asymptotic estimator of $\hat{\theta}$ [14, Theorem 2.2]. This is a significant results, as even in the absence of true model parameters θ_0, the MLE estimator asymptotically converges to the model parameters $\hat{\theta}$ that minimize the Kullback-Leibler divergence, thereby making $\hat{\theta}$ a well-defined object that we may seek to compute.

The asymptotic normality of the MLE estimator under misspecification can also be proved [14, Theorem 3.2]. Let us consider a sequence $Y_{\mathrm{exp},1}, \ldots, Y_{\mathrm{exp},N}$ iid according to $Y_{\mathrm{exp}}(x_{\mathrm{exp}}) = (Y_{\mathrm{exp}}(x_1), \ldots, Y_{\mathrm{exp}}(x_m))$. Then, the distribution of $\sqrt{N}(\theta_N^* - \hat{\theta})$ converges to $\mathcal{N}(0, C^{-1}(\hat{\theta}))$. The matrix C is akin to the Fisher information matrix (2.43), with slight modifications to take into account the fact that the model is now misspecified.

The asymptotic consistency and normality of the MLE estimator for a misspecified model are important. Even if the model h is not perfect, the use of MLE for the calibration should return the value of the model parameters that best fit the data in the sense of the Kullback-Leibler divergence (2.48). We will revisit these results in the context of goal-oriented calibration in Chapter 4.

The Bayesian calibration has also been studied for misspecified models. Without delving into the (very) technical details, it suffices to say that misspecified versions of the Bernstein-Von-Mises can be obtained, as shown in [15]. The posterior distribution asymptotically converges to a normal distribution centered around $\hat{\theta}$, the point that minimizes the Kullback-Leibler divergence (2.48), again with a covariance matrix akin to the Fisher information matrix.

Representation of the Modeling Errors

Several methods have been developed in the last decades to explicitly approximate the modeling errors. Such representation of the modeling errors can provide useful information re-

garding the ability of a model to capture reality. One of the first methods was proposed by Kennedy and O'Hagan in the early 2000's [18]. We will refer to this approach simply as the KOH framework. The method consists in enhancing the statistical model (2.37) with an additional Gaussian process E such that

$$Y(x) = h(x, \theta) + E(x) + \Phi(x). \tag{2.50}$$

The Gaussian process E represents the modeling discrepancy associated with model h. The precise nature of the modeling discrepancy is however unknown, so we want our representation E to be flexible. This is exactly the rationale of employing the KOH framework: a Gaussian process can represent a variety of functions, whose properties are completely defined by the kernel function.

The modeling discrepancy $E \sim \mathcal{GP}(m(x;\xi_m), k(x,x';\xi_k))$ is described by the set of parameters $\xi = (\xi_m, \xi_k) \in \mathcal{E} \subseteq \mathbb{R}^{d_E}$. They respectively define the mean and covariance function of the Gaussian process. Again, since the modeling discrepancy are not known a priori, we do not know the value of the parameters ξ. We then need to calibrate them, in conjunction with the model parameters θ. Several approaches can be adopted that either simultaneously or sequentially calibrate θ and ξ, as examined in [21].

With the modeling discrepancy E, we can analyze features of the modeling discrepancy, such as the regions in the parameter space $\mathcal{X}$ where the model performs well and the regions where the modeling discrepancy are large. It is important to mention that the KOH framework allows a representation of the modeling discrepancy affecting the observables of a system of interest. It does not provide information about the modeling discrepancy of the non-observable QoIs.

Another approach to estimate the modeling errors consists in changing the nature of the model parameters from deterministic to aleatory [25]. This modification is not made in the Bayesian sense, it is carried out by assuming that the model parameters follow a given distribution parametrized by the parameters α. For example, in [25], the authors assume that θ is assimilated as a polynomial chaos expansion described by α. These aleatory model parameters Θ can now capture some of the variability of Y_{exp}. Within this framework, we can rewrite relation (2.37) as

$$Y(x) = h(x, \Theta) + \Phi(x), \tag{2.51a}$$

$$Y(x, \omega) = h(x, \Theta(\omega)) + \Phi(x, \omega), \tag{2.51b}$$

where the second equation emphasizes the fact that a realization of Y is now obtained by

considering a realization of the model parameters Θ. In this way, the portion of the modeling errors that can be explained by the structure of the model h is accounted for via Θ. The calibration problem hence aims at determining the parameters α of the distribution of Θ. An important limitation of this approach is that the modeling errors that cannot be explained solely by the structure of the model h cannot be captured. However, the framework permits the propagation of some of the modeling errors to non-observable QoIs by evaluating $h_{\text{qoi}}(x, \Theta)$.

Another way to represent the modeling errors consists in recognizing that the model, encoded in the equilibrium relation (2.1), is incomplete and/or too coarse [26]. The idea is to incorporate a stochastic term to the equilibrium relation. For example, this could be done by modifying the original equilibrium relation (2.1) such that

$$\tilde{r}(x, \theta, A, \tilde{u}) = 0, \tag{2.52}$$

where $\tilde{r}$ represents the enhanced equilibrium relation, which includes additional physics encoded in the aleatory parameters A. The perturbed (and possibly expanded) state variable is denoted by $\tilde{u}$. The general representation (2.52) can consist in an additive or multiplicative perturbation of (2.1) or can even couple additional physics to the original model. Adopting this approach necessitates some additional knowledge on the system of interest at hand, as one needs to specify the relation $\tilde{r}$. Moreover, this approach is intrusive as solving the overall equilibrium relation (2.52) implies modifications of the original solver, the one employed to solve (2.1). The calibration process then calibrates the model parameters θ as well as the parameters describing the distribution of A. This methodology has the net advantage that the propagation of the modeling errors to the QoI is again possible and straightforward. We mention that a simpler way to include a stochastic term to the equilibrium relation that does not necessitate additional physical considerations consists in making the model parameters aleatory, as in Equation (2.51) [26, Section 4.2].

In Chapter 4, we will adopt the KOH framework to represent the modeling errors. We justify this choice by the necessity to have an explicit representation of the modeling discrepancy that does not depend on the model and the system of interest under study. With the approach of Sargsyan et al. [25], we can enhance the model by considering it stochastic (via the aleatory model parameters Θ). However, the method does not provide an explicit approximation of the modeling errors. The approach of Oliver et al. [26] provides an explicit approximation of the modeling errors by the relation $\tilde{r}$. However, the determination of $\tilde{r}$ is highly problem-dependent, which is cumbersome. We do not argue here that the KOH framework is the best way to represent modeling errors, only that for the perspective of developing a goal-oriented

calibration methodology, the KOH setting appears to allow the representation of the most general modeling discrepancy with respect to both x and θ.

2.8 Validation

> *All models are approximations. Assumptions, whether implied or clearly stated, are never exactly true. All models are wrong, but some models are useful. So the question you need to ask is not "Is the model true?" (it never is) but "Is the model good enough for this particular application?"*

This quote from Box et al. [82, p.61] encapsulates well the objective of validation. For the intended predictive objective of the model, one must assess whether the inevitable presence of modeling errors renders these predictions useless. If it is not the case, then one should ideally seek to quantify the confidence in the predictions. We present some approaches proposed to perform this validation task. We take inspiration from the presentation made in the article [33]. For the sake of the presentation, we will consider the model parameters as a random variable $\tilde{\Theta}$, whose distribution can be that of a posterior or that arising from considering the model parameters aleatory.

As argued in [33, Section 3], one can employ (binary) hypothesis testing to perform validation. More precisely, one encodes in a test statistic a feature of the experimental data deemed important with respect to assessing the validity of a model. A null hypothesis and an alternative hypothesis pertaining to this test statistic are formulated. Then, one computes the p-value of the experimental data, i.e., the probability of observing the sample assuming that the null-hypothesis holds. This p-value is compared against a predefined significance level and the null-hypothesis is rejected if the p-value is under this level. In the context of validation, the rejection of the null-hypothesis invalidates the model, whereas acceptance of the null-hypothesis indicates that the model is not invalidated. For the validation of physical models, it is not always straightforward to define a meaningful statistic, e.g. should one focus on specific moments of the distribution, should one perform a likelihood-ratio test, or should one compare the empirical distribution with that of the model. Moreover, even if a meaningful statistic were to be defined, its distribution can be difficult to obtain. For these reasons, we do not adopt this approach for the validation process. We refer the interested reader to the books of Schervish [12] and Lehmann and Romano [83] for a thorough presentation of statistical tests.

Validation Metric

Another approach to perform validation consists in quantifying the difference between the experimental data and the model predictions via validation metrics. By considering the predictive setting, the modeler/user can define an acceptable tolerance ε on the errors produced by the model. The validation metric should indicate whether this tolerance is respected or not.

One such validation metric is the area validation metric introduced by Ferson et al. [30] and employed by Roy and Oberkampf [29]

$$d(F_{Y_{\text{exp}}(x)}, F_{h(x,\tilde{\Theta})}) = \int \left| F_{Y_{\text{exp}}(x)}(y) - F_{h(x,\tilde{\Theta})}(y) \right| \mathrm{d}y. \tag{2.53}$$

This metric compares the cumulative distribution function of the true process Y_{exp} with the cumulative distribution function of the model observable prediction $h(x, \tilde{\Theta})$. An important feature of the area validation metric is that the distance d possesses the same (physical) unit as the observable. The value of $d(F_{Y_{\text{exp}}(x)}, F_{h(x,\tilde{\Theta})})$ can then be directly compared to the specified tolerance ε to assess the validity of the model.

Another validation metric is the model reliability metric introduced by Rebba and Mahadevan [31]

$$\gamma = \Pr\left(\left| Y_{\text{exp}}(x) - h(x, \tilde{\Theta}) \right| < \varepsilon \right). \tag{2.54}$$

This validation metric computes the probability γ that the difference between the observables and the model observables is less than the specified tolerance on the error ε. The value of γ hence provides a measure of the reliability of the predictions.

Extensions and improvements of the area validation and model reliability metric have been proposed in [84]. In Chapter 3, we will employ validation metrics to perform the validation task.

Model Selection

With the aforementioned validation metrics, multiple models may be deemed valid (not invalid). One then needs an additional criterion to determine which valid model to employ. This criterion is usually related to the Occam razor, i.e., amongst all plausible models, choose the most simple one. Let us suppose we possess a collection of N_M competing models $\mathcal{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_{N_M}\}$. In the following discussion, index i will refer to the i-th model. Two related ways of encoding the Occam razor criterion is via the Akaike Information Criterion

(AIC) or the Bayesian Information Criterion (BIC)

$$\mathrm{AIC}_i = -2 \ln f_{Y|\theta_i^*}(y_{\mathrm{exp}}) + 2k_i, \tag{2.55}$$

$$\mathrm{BIC}_i = -2 \ln f_{Y|\theta_i^*}(y_{\mathrm{exp}}) + k_i \ln N \tag{2.56}$$

where θ_i^* is the MLE solution for the i-th model, k_i is the number of components of θ_i, and N is the number of observations. We observe that the capacity of the model to capture the experimental data, encoded by the log-likelihood, is compensated by the complexity of the model, either via $2k_i$ or $k_i \ln(N)$. The best model is the one with the smallest AIC and/or BIC. A more detailed discussion regarding AIC and BIC and their relation to information theory can be found in [85].

Another way to compare models is via Bayes factors $B_{i,j}$ [33, 34]. Two competing models are compared according to

$$B_{i,j} = \frac{f_{Y|\mathcal{M}_i}(y_{\mathrm{exp}})}{f_{Y|\mathcal{M}_j}(y_{\mathrm{exp}})} = \frac{\int f_{Y|\Theta_i=\theta_i',\mathcal{M}_i}(y_{\mathrm{exp}}) f_{\Theta_i|\mathcal{M}_i}(\theta_i')\,\mathrm{d}\theta_i'}{\int f_{Y|\Theta_j=\theta_j',\mathcal{M}_j}(y_{\mathrm{exp}}) f_{\Theta_j|\mathcal{M}_j}(\theta_j')\,\mathrm{d}\theta_j'}, \tag{2.57}$$

where the numerator and denominator are the (model) evidence for model i and j, respectively. A Bayes factor $B_{i,j} \ll 1$ indicates that the model j should be favored, whereas $B_{i,j} \gg 1$ indicates that the model i better explains the experimental data.

Closely related to the Bayes factors are the posterior model plausibility ρ_i [35–37]

$$\rho_i = \frac{f_{Y|\mathcal{M}_i}(y_{\mathrm{exp}}) \Pr(\mathcal{M}_i|\mathcal{M})}{\sum_j^{N_M} f_{Y|\mathcal{M}_j}(y_{\mathrm{exp}}) \Pr(\mathcal{M}_j|\mathcal{M})}, \tag{2.58}$$

where $\Pr(\mathcal{M}_i|\mathcal{M})$ is the prior of model $\mathcal{M}_i$ given the set of candidate models $\mathcal{M}$. The normalization ensures that $\rho_i \in (0, 1)$. From (2.58), we can observe that ρ_i is related to the posterior probability of model $\mathcal{M}_i$. Hence, if all models of the set $\mathcal{M}$ are valid, the most probable one (the one with the largest ρ_i) should be chosen.

A systematic procedure referred to as the Occam-Plausibility Algorithm (OPAL) describes a way to select the simplest valid model [36–38]. A set of candidate models are separated according to their complexity, usually measured by the number of components of the model parameters θ_i. The models with the lowest complexity are calibrated with experimental data and the model with the highest posterior model plausibility (2.58) is retained. This model is then validated according to a validation metric. If it is validated, then this model is the simplest plausible model. If not, the procedure is repeated for models of greater complexity.

Validation of QoIs

All the aforementioned validation methods assess the validity of a model with respect to the prediction of observables. However, they do not directly indicate if the model is valid to perform predictions of non-observable QoIs. Our point of view is that such direct assessment of the validity of prediction of non-observable QoIs will always remain elusive. We must then extrapolate the validation assessment of the predictions of observables onto that of the QoIs. The challenge is to ensure that there is a correspondence between the observables and the QoIs so that the extrapolation can be deemed meaningful. We tackle this specific problem in Chapter 3.

As a final note, we mention the works of Oden and Prudhomme [39] and of Morrison et al. [40]. These works study the sensibility of the QoI to the calibration process. More precisely, a first calibration is performed with calibration data and a distribution for the QoI is obtained. Then, a second calibration is performed with the calibration and validation data and a second distribution of the QoI is obtained. The two distributions of the QoI are then compared. If the distance is large, it indicates that the prediction of the QoI is sensitive with respect to the calibration. It is not a validation process *per se*, but the methodology can still provide valuable information in that one should be careful in predicting QoIs that are sensitive to the calibration process.

CHAPTER 3 OPTIMAL DESIGN OF VALIDATION EXPERIMENTS FOR THE PREDICTION OF QUANTITIES OF INTEREST

This chapter is based on the paper *Optimal design of validation experiments for the prediction of quantities of interest* [41], published in the journal *Computer Methods in Applied Mechanics and Engineering* in 2023 and co-authored with Serge Prudhomme and Marc Laforest. We have seen in Section 2.8 that any validation process relies on experimental data gathered from a system of interest at given scenarios. Moreover, the validation assessment is usually performed with respect to the observables rather than the QoIs.

In this chapter, we address the issue of the design of experiments employed for validation. More specifically, we want to design a validation experiment that is tailored toward assessing the validity of a non-observable QoI. The first issue tackled in this chapter consists in determining an appropriate validation scenario in cases where the prediction scenario cannot be carried out in a controlled environment. The second issue concerns the selection of observations when the quantity of interest cannot be readily observed. The proposed methodology involves the computation of influence matrices that characterize the response surface of given model functionals. Minimization of the distance between influence matrices allows one to select a validation experiment most representative of the prediction scenario. We illustrate our approach on two numerical examples. The first example considers the validation of a simple model based on an ordinary differential equation governing an object in free fall to put in evidence the importance of the choice of the validation experiment. The second numerical experiment focuses on the transport of a pollutant and demonstrates the impact that the choice of the quantity of interest has on the validation experiment to be performed.

3.1 Introduction

The development of models to describe a system of interest (be it physical, biological, or societal) is central in understanding and predicting its behavior. Mathematical models may stem from physical laws and/or empirical considerations for either some parts of the system or the whole system itself. In order to assess the potential of a model to explain and predict a given quantity of interest (QoI), it must be validated, or at best, not invalidated. In other words, the error between the model and the reality it is supposed to describe must be precisely quantified with respect to the QoI. The validation process implies comparing outputs of the model with experimental data acquired by testing the system of interest. Over the

past decades, several procedures for model validation have been proposed in the literature. For instance, Oberkampf and Trucano [28] present a detailed overview of verification and validation processes while Roy and Oberkampf [29] provide a thorough discussion on these two processes with a focus on the treatment of both aleatory and epistemic uncertainty. A recent review article by Riedmaier et al. [86] discusses in details the scope and shortfalls of existing approaches for validation and prediction. The paper of Lee et al. [87] provides an extensive review of the literature addressing the problems of uncertainty propagation, parameters calibration, and model validation. Model validation has become increasingly relevant and primordial in the field of computational sciences and engineering due to the complexity of scientific models currently employed and the demand for predictive and quantitative results. However, a critical aspect remains the selection of suitable validation experiments, since they provide the experimental data against which the model prediction is compared and thus directly influence the decision whether the model is deemed valid or not. This issue is even more important when the number of validation experiments or amount of data are limited. Yet, the precise design of such validation experiments is still a question largely overlooked in the literature, with the exception of a few contributions.

Oliver et al. [26] stress the importance of the validation process for predictive modeling, even more so in the case of models made of several coupled sub-models. They essentially identify two fundamental issues when tackling model validation: 1) that of validating a model if one cannot reproduce the conditions under which the predictions are made; 2) that of validating a model if the QoI one wishes to predict cannot be observed in practice. To answer these two questions, the authors estimate the modeling errors affecting the sub-models and propagate them to the QoI at the prediction scenario. They then propose a series of guidelines to ensure that the calibration and validation experiments are relevant for prediction purposes [26, Section 3.3]. These guidelines can be roughly summarized as follows: if the QoI is sensitive to certain model parameters and/or certain modeling errors, then the calibration and validation experiments should reflect these sensitivities. However, the discussion remains essentially qualitative and the analysis is performed a posteriori, that is, once calibration and validation experiments have been performed, they subsequently verify that these are actually relevant.

The paper of Hamilton and Hills [88] and the subsequent report of Hills [89] also address these aforementioned issues. The authors develop a meta-model allowing the extrapolation of the modeling error to the QoI at prediction scenario. More precisely, they identify a linear relationship between the sensitivities (some kind of elementary effects) of the QoI at prediction scenario to the sensitivities of the observables at validation scenario. Weights could then be attributed to more relevant validation observations with respect to the objective of the

prediction. However, the assessment is again made a posteriori. Moreover, the authors do not address the design of validation experiments. We nevertheless note that Hills [89, Section 1.3] presents several cases where a validation experiment may fail to provide a useful assessment on whether the model can be employed to predict a QoI at a prediction scenario.

Other works have also studied the problem of validating mathematical models made of multiple sub-models. Among those, we mention the work of Hills and Leslie [90], who employ local derivative-based indices to weight the importance of the validation experiments for the sub-models with respect to the numerical prediction using the full model. In the same vein, the work of Li and Mahadevan [91] proposes an integrated approach to take into account the validity of the lower-level models and their relevance for the system-level model. More precisely, they employ a model reliability metric to assess the validity of lower-level models and perform a sensitivity analysis based on Sobol indices [57] to quantify the relation of these same lower-level models to the model for the full system.

Farrell et al. [36] propose the so-called Occam-Plausibility algorithm that allows one to identify the simplest and most plausible model among a given family of competing models that they rank in terms of the posterior model plausibility. The paper of Tan et al. [38] enriches the Occam-Plausibility algorithm by providing two criteria about the selection of the validation experiment to be performed. The first criterion stipulates that the sensitivity indices of the validation scenario must be close to the sensitivity indices of the prediction scenario. They employ Sobol indices [57] as their sensitivity indices, thus considering the decomposition of the variance as a notion of sensitivity. This criterion is also verified a posteriori, once a validation scenario has been proposed. The second criterion checks if the validation scenario provides an information gain on the posterior distribution of the model parameters compared to the calibration scenarios.

Ao et al. [92] specifically study the optimal design of validation experiments in the context of life prediction models. Their approach is actually similar to the Bayesian optimal design approach described in Section 3.2.6. However, their methodology relies on a fine model representing the system of interest, thus necessitating additional a priori knowledge on the system of interest, which is not always available.

We also mention the paper of Sunseri et al. [93] that investigates the influence of auxiliary parameters on the solution of inverse problems. Although their study does not address the problem of validation, their methodology is relevant to the present work. They determine some pointwise and global sensitivity indices using the gradient of the solution to the inverse problem with respect to the auxiliary variables. Whenever an auxiliary parameter possesses a high sensitivity index, it indicates that a small perturbation in its value significantly impacts

the solution to the inverse problem and that some effort should be spent in order to decrease the uncertainty of this auxiliary parameter.

The present work aims at refining some of the concepts laid out in the aforementioned papers, in particular, the use of sensitivity indices to assess the relevance of validation experiments. The main objective here is therefore to describe a novel methodology for the optimal design of validation experiments. As physical models are becoming increasingly more complex and are being applied to a larger range of critical applications, it is thus important to ponder each step of the validation process. To the best of the authors' knowledge, the design of validation experiments is generally guided by expert opinions or sometimes relies on a heuristic process. The proposed methodology provides a systematic way of designing a validation experiment, even if the QoI is not observable or the prediction scenario is not experimentally producible. Moreover, it does not require any experimental data. The approach essentially relies on the formulation of two distinct optimization problems, which are formulated in such a way that the behavior of the model under the validation conditions resembles as much as possible that of the same model, but under the prediction conditions.

The paper is organized as follows. Section 3.2 introduces all the necessary concepts and notations to precisely describe the validation process. A detailed account of the different types of parameters and sources of uncertainty involved in a model will be provided. We also describe methodologies for calibration and validation and give a brief overview of methods for optimal design of calibration experiments. Section 3.3 presents our novel approach for the design of validation experiments tailored toward the prediction of a quantity of interest that cannot potentially be observed. We briefly review the Active Subspace method as the chosen method to perform the sensitivity analysis. Two optimal design problems will be formulated in order to compute the appropriate control and sensor parameters for the validation scenario. We then present some numerical examples in Section 3.4. We first consider a toy problem, that of an object in free fall, to illustrate the validation process, with an emphasis on the importance of designing optimal validation experiments. We then study a second example, consisting of a simplified pollutant transport problem, to highlight the importance of the definition of the QoI in the design of optimal validation experiments. Section 3.5 provides some concluding remarks along with a series of open questions regarding the proposed methodology and how it relates to the larger process of uncertainty quantification and model validation.

3.2 Terminology and Preliminary Notions about Validation and Predictive Modeling

The purpose of this section is to introduce some preliminary notions and notations in order to precisely describe the validation process. We recall that the primary objective of predictive modeling is to be able to obtain some quantitative predictions regarding a system of interest. At first sight, the task may seem rather straightforward to achieve as it simply aims at evaluating the output of a model. However, as mentioned earlier, for numerical predictions to be of any usefulness, one should be confident that the model remains valid for its intended use and that the predictions obtained with the model are in fact accurate descriptions of the reality. It should be clear by now that the validity of the model should be assessed with respect to the quantities of interest that one would like to predict using the model at specific conditions/regimes. In the presentation below, we will strive to use the same terminology as that commonly introduced in the literature, with a few exceptions in order to emphasize the different roles of some of the model parameters and model outputs.

3.2.1 Abstract Model

We define a generic *deterministic model* as the triplet (r, p, u), where $p = (p_1, \ldots, p_d) \in \mathcal{P}$ are the *parameters*, $u = (u_1, \ldots, u_s) \in \mathcal{U}$ are the *state variables*, and r include the *equilibrium relation* and initial and/or boundary conditions

$$r(p, u) = 0. \tag{3.1}$$

For the sake of simplicity, the parameter space $\mathcal{P} \subseteq \mathbb{R}^d$ is assumed to be finite dimensional and the state space $\mathcal{U}$ to be either a Banach or Hilbert space. The equilibrium relation r may involve ordinary or partial differential equations, algebraic equations, or any combination thereof. The equilibrium relation encapsulates the set of hypotheses and scientific understanding that hopefully lead to a sufficiently accurate model of the system of interest. An important property that the equilibrium relation r must possess is well-posedness: for each $p \in \mathcal{P}$, there exists a unique and stable solution $u \in \mathcal{U}$ satisfying (3.1). In other words, there exists a function g such that $u = g(p)$.

As an example, we consider a simple model that will be used in Section 3.4.1 to illustrate our methodology. The model describes the motion of a spherical projectile launched vertically in ambient air at sea level. The evolution of the physical system can be described by the linear

ordinary differential equation and initial conditions

$$m\frac{\mathrm{d}^2u}{\mathrm{d}t^2}(t) + 3\pi\ell\exp(\mu)\frac{\mathrm{d}u}{\mathrm{d}t}(t) = -mg, \qquad \forall t \in (0,T) \tag{3.2a}$$

$$u(0) = u_0, \tag{3.2b}$$

$$\frac{\mathrm{d}u}{\mathrm{d}t}(0) = v_0, \tag{3.2c}$$

where $u(t)$ is the altitude reached by the projectile at time t, m and ℓ are its mass and diameter, u_0 and v_0 are its initial position and velocity, respectively, g is the gravitational constant, and μ is a viscosity parameter for a spherical object travelling in air [94, p.488]. In this example, the parameters of the problem are given by $p = (m, \ell, u_0, v_0, g, \mu) \in \mathcal{P} = (\mathbb{R}^+)^5 \times \mathbb{R}$ and the state variable consists in the altitude $u \in \mathcal{U} = C^1(\mathbb{R})$ equipped with the uniform norm on u and $\mathrm{d}u/\mathrm{d}t$. The set of equilibrium relation and initial conditions can easily be obtained from the initial-value problem (3.2). The model is constructed by making the following hypotheses regarding the physical system:

1. The trajectory of the projectile is supposed to be one dimensional, so only forces acting vertically are taken into account.
2. The friction force is assumed to be proportional to the velocity of the projectile.
3. The viscosity of the air and the gravitational constant remain constant with respect to the altitude.

3.2.2 Classes of Parameters

We split the parameters p of a problem into two distinct subsets, namely the *control parameters* x and the *model parameters* θ. We also consider an additional set of parameters that we call the *sensor parameters* $z \in \Omega$, where Ω is the domain of definition of the state variable u. Each type of parameters will play some specific roles, as described below. Similar definitions can be found in the papers cited in the introduction, even if no consensus has yet been reached in the literature on a common terminology.

The *control parameters* $x \in \mathcal{X}$ represent the parameters that one can control when performing an experiment on the system of interest. These parameters can be viewed as the set of inputs which uniquely determine the experimental scenario. Examples of such parameters are the parameters appearing in the initial conditions and boundary conditions, the geometrical data, and the physical properties directly measurable by an experimental apparatus (mass, length, temperature, etc.). For the example of Section 3.2.1, the control parameters are given by the mass and diameter of the projectile as well as its initial altitude and velocity,

hence $x = (m, \ell, u_0, v_0)$.

The *model parameters* $\theta \in \mathcal{T}$ are the remaining parameters necessary for the full description of the model ($\theta = p \setminus x$). They are coined this way due to the fact that they explicitly stem from the assumptions and simplifications that one makes in defining a model to describe a particular phenomenon. Other hypotheses could potentially lead to a new model with different model parameters, but will not change the control parameters. Naturally, the definition of the parameters p, and hence of the model parameters θ, depends directly on what a parameter would be deemed as such by the modeler and/or user. For example, numerical constants like π in (3.2) clearly do not constitute parameters, but the picture could be far less clear for other quantities. As a rule of thumb, if the value of a quantity may vary or if it is of any interest in the analysis of the model, then it can be considered as a model parameter. Examples of model parameters can be those that characterize the properties of a material (e.g. Young modulus, Poisson coefficient). Model parameters are specified either by prior knowledge or by a calibration process. A more detailed discussion about the latter point will be presented in Section 3.2.5. In the case of the spherical projectile, the model parameters are $\theta = (g, \mu)$ since they result from assumptions in the modeling process (g is independent of the altitude and μ is related to a sub-model for air friction).

A final type of parameters are the so-called *sensor parameters* $z \in \Omega$. They often represent location defined in terms of the independent variables (e.g. space or time) upon which the state variable u would be observed. In the example of the projectile, the sensor parameters will be the time at which we wish to evaluate the altitude, so $z = t$.

3.2.3 Classes of Scenarios

The control and model parameters allow one to define the regime under which both the system and the model operate. A specific regime will be referred to as a *scenario*. For the sake of simplicity here, we shall restrict the model parameters θ to be the same for the validation and prediction scenarios. For the projectile example, this implies that the prediction and experimental apparatus be at sea level and in the ambient atmosphere (same gravitational constant g and same viscosity μ). In other words, only the control parameters x are needed to fully characterize a scenario. We shall study three classes of scenarios, with a particular emphasis on the first two.

1. **Prediction Scenario:** Defines the conditions, described by x_{pred}, under which predictions on the system of interest will be made.
2. **Validation Scenario:** Defines the conditions, described by x_{val}, under which the vali-

dation experiment will be carried out.

3. **Calibration Scenario:** Defines the conditions, described by x_{cal}, under which the calibration experiment will be performed. We note here that model calibration essentially aims at estimating the uncertainty of the model parameters θ.

On the one hand, the validation and calibration experiments are performed within a *set* $\mathcal{X}_{\text{lab}} \subset \mathcal{X}$ of *controlled environments*. On the other hand, the prediction scenario x_{pred} may or may not be producible experimentally, so that x_{pred} may not belong to $\mathcal{X}_{\text{lab}}$. Figure 3.1 provides a conceptual illustration of the various scenarios.

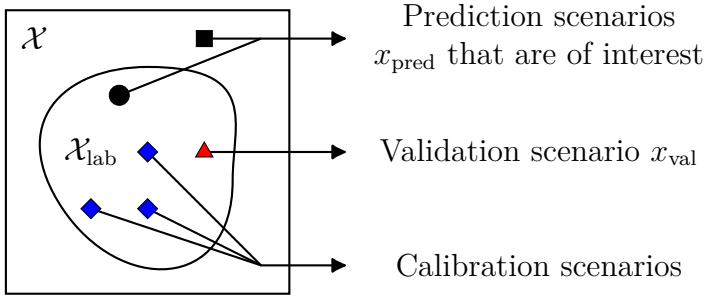


Figure 3.1 Illustration of the various scenarios. The prediction scenario may be producible (black circle) or may not be producible experimentally (black square). The validation and calibration scenarios are elements of the set $\mathcal{X}_{\text{lab}}$ of controlled environments.

3.2.4 Observables and Quantities of Interest

We reiterate that a mathematical model constitutes only an abstraction of a physical system of interest. In fact, one has access to limited information about the system of interest. This information will be collectively referred to as the *experimental observables* y_{exp}. The experimental observables y_{exp} can be acquired through a variety of experimental devices and/or statistical surveys. Importantly, they are obtained for a particular scenario and for given sensor parameters, so that they are functions of x and z, i.e., $y_{\text{exp}} = y_{\text{exp}}(x, z)$. We also assume that the model (r, p, u) is sufficiently rich so that each experimental observable can be abstracted by a functional of the parameters p and the sensor parameters z (the state variable u is implicitly given through the function $g(p)$). These observables will be referred to as the *model observations* $y \coloneqq h_{\text{obs}}(p, u, z) = h_{\text{obs}}(p, z)$. In the following, the term *observables* will be used to refer to both experimental and model observations. In the example dealing with the launch of a spherical projectile, the experimental observables may consist in the altitude of the projectile, acquired by an optical apparatus or an altimeter, or the acceleration of the projectile obtained with an accelerometer. Whatever the quantity

considered, it will be measured at a given discrete time $z = t_{\text{exp}}$. The corresponding model observations are thus defined as $y(t_{\text{exp}}) \coloneqq u(t_{\text{exp}})$ or $y(t_{\text{exp}}) \coloneqq \mathrm{d}^2u/\mathrm{d}t^2(t_{\text{exp}})$.

Quantities of interest extend the notion of model observables, in the sense that they refer to quantities that can or cannot be experimentally observed. In other words, QoIs may not depend on sensor positions, although they could depend implicitly on the independent (spatial or temporal) variables through an average or a maximum. The QoI will be denoted as $q \coloneqq h_{\text{qoi}}(p, u) = h_{\text{qoi}}(p)$, since they are evaluated for specific scenarios. Similarly to h_{obs}, h_{qoi} will be considered a functional. We list below some examples of such possible QoIs:

- Mean quantities over a subregion of Ω (e.g. mean displacement or stress in a solid);
- Minimal or maximal quantities and their location (e.g. maximal stress);
- Statistical quantities (e.g. probability of leakage of nuclear waste).

An important notion is that the QoIs provide an abstraction of the key features of the system of interest upon which a decision making process will be applied. Figure 3.2 synthesizes how the observations and QoIs are obtained in terms of the various parameters. In the case of the projectile, an example of QoI could be the maximal altitude that it reaches over time, i.e., $q \coloneqq \max_t u(t)$. We note that the quantity is not directly observable since we would need to know beforehand the exact time at which the maximal altitude of the real projectile is reached.

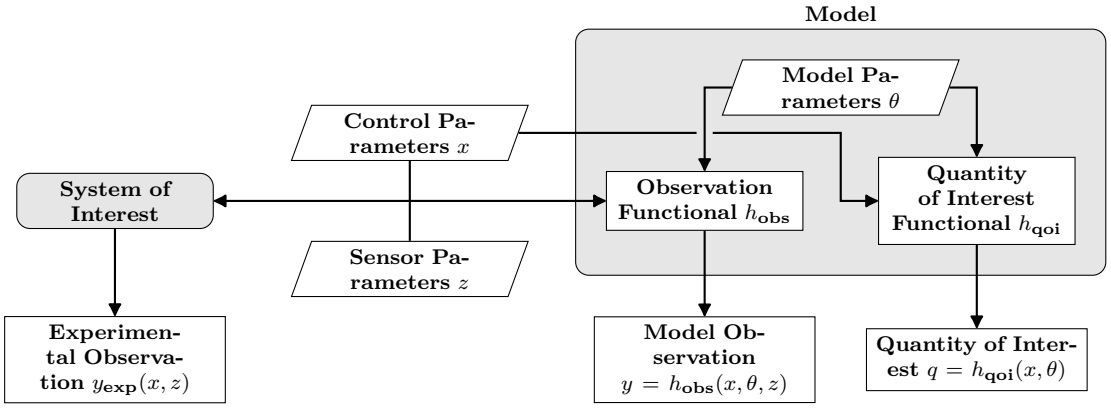


Figure 3.2 Diagram of the interactions between the various parameters and the system of interest as well as the model. It is important to observe that the model parameters θ do not affect the system of interest, but only the observation and QoI functional h_{obs} and h_{qoi}. For the sake of simplicity, the QoI functional h_{qoi} does not depend explicitly on the sensor parameters z.

A precise description of the prediction problem can now be stated as follows: one wishes to predict a given QoI h_{qoi} for a given prediction scenario x_{pred}. The definition of observable

functionals h_{obs} will play a crucial role in the calibration and validation problems in order to obtain accurate predictions. The accuracy of the model outputs (observables and QoI) also depends upon the aggregation of various sources of uncertainties and errors, as explained in the following subsection.

3.2.5 Uncertainties and Modeling Errors

Several approaches can be adopted to represent uncertainties in quantities, such as parameters, observables, or QoI, as outlined in [43, 86]. We consider here the probabilistic approach, where an uncertain quantity is assimilated as a random variable A with probability distribution F_A and density function f_A. The random variable A can be either discrete or continuous. We adopt the usual convention that a lowercase variable a indicates one realization of the random variable A, noted in uppercase. In addition to random variables, we may also encounter random processes $A = A(b)$, usually consisting of a family of random variables, each defined for an instance of b.

Uncertainty in the Parameters

As a first approximation, we suppose that the control parameters x and the sensor parameters z do not possess any uncertainty, an assumption that can be far from the truth depending on the nature of the prediction scenario and of the actual experimental apparatus. However, this hypothesis will simplify the definition of the optimal validation experiment as well as the overall exposition of the present work. Therefore, the only parameters that we will be considered uncertain here are the model parameters θ.

The quantification of the uncertainty in the model parameters θ is a non-trivial task since this type of parameters arises from the various hypotheses of the model. One can determine the distribution F_Θ by taking into account prior knowledge concerning each hypothesis. In the example of the projectile, we possess relatively extensive prior knowledge on the gravitational constant g and the air viscosity μ, so one can assign them a distribution with small variance. However, it is often the case that we have a rather vague idea about the values of the model parameters. The probability distribution of each parameter need then be identified through calibration processes. This particular topic is the subject of an exhaustive literature that will be briefly covered in Section 3.2.6. For the sake of simplicity, we assume that the uncertainties affecting the model parameters θ are relatively well-known (whether by prior knowledge or by a calibration process) so that a probability distribution F_Θ can be prescribed. Moreover, we suppose that the model parameters are identifiable and that they have indeed been identified. These simplifying assumptions allow us to avoid delving into other issues,

such as identifiability and the distinction between aleatory and epistemic uncertainties.

Uncertainty in the Observables and QoIs

Uncertainty in experimental observations may come from various sources and may depend on the scenarios x and sensor parameters z. Experimental observations are thus assimilated as random processes $Y_{\text{exp}}(x, z)$. Being able to characterize the uncertainty in the observables is crucial for the validation process, as further explained in Section 3.2.7. We list two possible, although not exhaustive, sources of uncertainty. The first one comes from the inherent and inevitable variability of the measurement apparatus. However, it is sometimes possible to reduce the amount of uncertainty in the observables and thus get experimental observations closer to the *true* value by using a more precise apparatus. The second source of uncertainty comes from the inherent stochastic nature of the system of interest. For fixed control and sensor parameters x and z, repetitions of the same experiment may yield different, yet accurate, values for the experimental observations. This may happen for instance if the phenomena of interest suffer from instabilities with respect to infinitesimal perturbations of the control parameters. The aggregation of these two sources of uncertainty allows one to determine the distribution $F_{Y_{\text{exp}}(x,z)}$, but the task is not necessarily straightforward. We take the point of view that we can repeat a certain number of times the same experiments (in the case of non-destructive experiments) so that one can determine the empirical distribution $F_{Y_{\text{exp}}(x,z)}$, which aggregates both sources of uncertainty.

Uncertainty in the model observation Y and quantity of interest Q are obtained by propagating the uncertainty in the model parameters Θ through their respective functional

$$Y(x, z) \coloneqq h_{\text{obs}}(x, \Theta, z), \tag{3.3}$$

$$Q(x) \coloneqq h_{\text{qoi}}(x, \Theta). \tag{3.4}$$

Computing the exact distributions F_Y and F_Q is only possible for very simple functionals and for a restricted family of distributions F_Θ (e.g. Gaussian distributions). Most of the time, one needs to approximate the distributions F_Y and F_Q by sampling the distribution F_Θ using methods such as the ubiquitous Monte-Carlo method, the Latin Hypercube method, or Quasi Monte-Carlo methods [3]. It is often enlightening to perform an uncertainty propagation prior to any validation step. If the prediction of the Q possesses too much uncertainty, then the prediction may be useless, independently of whether the model is deemed valid or not. In this case, it may be necessary to refine the calibration of the model parameters Θ in order to reduce their uncertainty, thereby potentially reducing as well the uncertainty in the QoI.

The use of sensitivity analysis methods [50, 95] can help identify the model parameters Θ contributing the most or least to the uncertainty in Y and Q. We summarize the various quantities introduced so far as well as the source of their uncertainty in Table 3.1.

Modeling Errors

The authors adhere to the common idea that that no model, as sophisticated as it may be, would be able to capture the exact behavior of a system of interest, since a model is merely an abstraction of the latter and often relies on several hypotheses. Therefore any model comes with some modeling errors that should be estimated in one way or the other. Furthermore, the modeling error should account for the inevitable sources of uncertainty in both the experimental and model observations. One way to combine parameter uncertainties and modeling errors is in terms of the discrepancy E:

$$E(x, z) = Y_{\text{exp}}(x, z) - Y(x, z) = Y_{\text{exp}}(x, z) - h_{\text{obs}}(x, \Theta, z). \tag{3.5}$$

The discrepancy $E(x, z)$ can be viewed as a random process in the sense that it consists of a random variable for each instance of the control parameters x and of the sensor parameters z. This discrepancy E can be viewed as the aggregation of the experimental errors and the model inadequacy in the approach taken by Kennedy and O'Hagan [18, 96]. An important property of the discrepancy (3.5) is that it is defined independently of the QoI. Other approaches can be adopted to quantify the modeling errors [26, 89]. For the sake of simplicity in the presentation, and since our focus is rather on the design of validation experiments, we omit to describe these methods here. We refer the interested reader to [86, 97] for a summary of these methods.

3.2.6 Calibration Process and Optimal Design of Calibration Experiments

We provide here a brief overview of the calibration process and of the related problem dealing with the optimal design of calibration experiments. Although our proposed approach to perform optimal design of validation experiments does not directly involve a calibration process *per se*, a brief discussion on the topic is deemed insightful in order to differentiate our methodology from the optimal design problem for calibration experiments.

The calibration of the model parameters Θ, a process usually referred to as model calibration or parameter identification, has as a main objective to provide a characterization of the probability distribution F_Θ, which encapsulates the uncertainty in the model parameters, using actual experimental observables y_{exp}, as realizations of Y_{exp}. In other words, the goal

Table 3.1 Summary of the parameters, observables, and QoI as well as their source of uncertainty

	Definition	**Uncertainty**
Control Parameters x	Parameters that one can control to perform an experiment	None
Model Parameters Θ	Remaining parameters necessary to describe a model	Prior knowledge or posterior density
Sensor Parameters z	Spatio-temporal parameters	None
Experimental Observation $Y_{\text{exp}}(x, z)$	Observations at control parameters x and sensor parameters z	Empirical density
Model Observation $Y(x, z)$	$Y(x, z) \coloneqq h_{\text{obs}}(x, \Theta, z)$	Uncertainty propagation
Quantity of Interest $Q(x)$	$Q(x) \coloneqq h_{\text{qoi}}(x, \Theta)$	Uncertainty propagation

is to update our knowledge about the model parameters. There exist several methods to perform a calibration, for example, the least squares approach (and its regularized version), the maximum likelihood estimation (MLE) approach, or the maximum a posteriori (MAP) approach. Under various hypotheses, one can show equivalences between these approaches [9]. A common feature of these methods is that they provide a point estimator $\hat{\theta}$ of the model parameters θ. These approaches present two drawbacks that are closely related. First, it is relatively difficult to quantify the uncertainty of $\hat{\Theta}$ and only asymptotic results may be applicable [13, Section 16.4]. Second, if the model is not *identifiable*, then these asymptotic results cannot be invoked and the point estimator $\hat{\theta}$ may not be "stable". This identifiability condition is a restrictive condition to fulfil and especially tedious to verify [81].

An alternative approach consists in updating the uncertainty of the model parameters with the posterior of a Bayesian analysis. As for the MLE and MAP approaches, the method requires defining a likelihood function $L_{Y_{\text{exp}}=y_{\text{exp}},x_{\text{cal}},z_{\text{cal}}}(\theta)$, which describes the probability of obtaining the experimental observables $Y_{\text{exp}} = y_{\text{exp}}$ given the control, sensor, and model parameters, x_{cal}, z_{cal}, and θ, respectively. In addition to the likelihood, one must provide a prior density, denoted by f_{Θ_0}, encoding the current knowledge that one possesses about the model parameters Θ. With the likelihood function and the prior density, Bayes' theorem allows one to obtain the posterior probability density function $f_{\Theta|Y_{\text{exp}}=y_{\text{exp}},x_{\text{cal}},z_{\text{cal}}}$ as

$$f_{\Theta|Y_{\text{exp}}=y_{\text{exp}},x_{\text{cal}},z_{\text{cal}}}(\theta) = \frac{L_{Y_{\text{exp}}=y_{\text{exp}},x_{\text{cal}},z_{\text{cal}}}(\theta) f_{\Theta_0}(\theta)}{\int L_{Y_{\text{exp}}=y_{\text{exp}},x_{\text{cal}},z_{\text{cal}}}(\theta) f_{\Theta_0}(\theta) \mathrm{d}\theta}. \tag{3.6}$$

The posterior density represents the updated knowledge about the model parameters Θ

given the experimental observations $y_{\text{exp}}(x_{\text{cal}}, z_{\text{cal}})$. The analytical expression of the posterior density is available only in a few cases, for example, when the model is linear with respect to the model parameters and one chooses a Gaussian likelihood and prior. Otherwise, it is generally approximated using sampling methods such as the Markov Chain Monte-Carlo method or its variants [3, 98].

Since all calibration approaches require the use of experimental observations $y_{\text{exp}}(x_{\text{cal}}, z_{\text{cal}})$, the calibration scenario x_{cal} and sensor parameters z_{cal} should be chosen appropriately in order to better determine the uncertainty in the model parameters Θ. To the best of our knowledge, there essentially exist two methods for optimal design of calibration scenarios, one based on the Fisher information matrix and the other based on a Bayesian approach.

1. The first method consists in optimizing some functional of the Fisher information matrix associated with the model [99,100]. This approach is widely used for polynomial models (linear with respect to the model parameters and possibly non-linear with respect to the control and sensor parameters) since the Fisher information matrix depends only upon the control and sensor parameters [99]. Different definitions of the Fisher information matrix functional lead to different optimal calibration experiments. For instance, the D-optimal design minimizes the variance of the estimator of the model parameters θ and the G-optimal design minimizes the maximum of the variance in the prediction. The D-optimal problem for a linear model reads

$$(x_{\text{cal}}, z_{\text{cal}}) = \underset{(x,z)\,\in\,\mathcal{X}_{\text{lab}}\times\mathcal{Z}_{\text{lab}}}{\operatorname{argmax}} |\mathcal{I}(x, z)|, \tag{3.7}$$

   where $\mathcal{I}$ denotes the Fisher information matrix. Equivalence theorems show that these different optimal designs are in fact related. For a thorough discussion of this approach and some extensions to non-linear models, we refer the interested reader to the book of Atkinson et al. [100].
2. The second method is based on the Bayesian approach [101–103] and focuses on minimizing some functional of the posterior $f_{\Theta|Y=y_{exp},x_{\text{cal}},z_{\text{cal}}}$ (or sometimes the likelihood function) over the possible observations Y and the possible values of the model parameters Θ. Definitions of the functional akin to the one employed with the Fisher information matrix give rise to similar optimal designs. For example, we can seek the control parameters x_{cal} and sensor parameters z_{cal} that minimize the determinant of the covariance of the posterior

$$(x_{\text{cal}}, z_{\text{cal}}) = \underset{(x,z)\,\in\,\mathcal{X}_{\text{lab}}\times\mathcal{Z}_{\text{lab}}}{\operatorname{argmax}} \iint \Big(\det(\operatorname{Cov}(\Theta|Y = y_{exp}, x, z))\Big)^{-1} \mathrm{d}F_{Y_{\text{exp}}}\, \mathrm{d}F_{\Theta_0}. \tag{3.8}$$

We remark that solving these types of optimization problems can be very challenging because it usually requires a nested integration. Several simplifications, such as a Laplace approximation of the posterior, can make these optimization problems tractable [101, 102].

A very important feature of the two approaches is that the determination of the optimal calibration experiment does not rely on any experimental data, as it is an a priori analysis. We will retain this feature in our methodology to design optimal validation scenarios.

3.2.7 Validation Process

Everything is now in place to discuss the problem of validating a model. Again, the goal is to assess the accuracy in the prediction of a quantity of interest $Q = h_{\text{qoi}}(x_{\text{pred}}, \Theta)$ at the prediction scenario x_{pred}. We reiterate that the discrepancy E given in (3.5) is a useful way to encode the accuracy of our model observations. Both experimental and model observations should be performed at a specific scenario, the so-called validation scenario x_{val}, and for given sensor parameters z_{val}. For now, let us suppose that we have determined this validation scenario x_{val}, the type of observable (i.e., the choice of the functional h_{obs}) and the value of the sensor parameters z_{val}. Multiple approaches to validate a model have been proposed, as mentioned in the introduction. Several of these validation processes require the choice of a validation metric. For example, Roy and Oberkampf [29] utilize the area validation metric introduced by Ferson et al. [30]

$$d(F_Y, F_{Y_{\text{exp}}}) = \int_{-\infty}^{\infty} \left| F_Y(s) - F_{Y_{\text{exp}}}(s) \right| \mathrm{d}s, \tag{3.9}$$

where F_Y and $F_{Y_{\text{exp}}}$ are the model and experimental observation distributions, respectively. If the value of $d(F_Y, F_{Y_{\text{exp}}})$ is superior than some tolerance d_{tol}, then the model is deemed invalid. They provide several rationales for this choice of validation metric and provide several ways to employ it, especially in the presence of different validation scenarios. Another validation metric employs the discrepancy E with a tolerance ε on the error

$$\gamma = F_{|E|}(\varepsilon) = \Pr(|E| < \varepsilon) = \Pr(|Y_{\text{exp}}(x_{\text{val}}, z) - h_{\text{obs}}(x_{\text{val}}, \Theta, z)| < \varepsilon). \tag{3.10}$$

If the value of γ is above a certain threshold η, then the model is deemed valid. This particular measure was introduced by Rebba and Mahadevan [31] and coined *model reliability metric*. In their paper, the authors compare the model reliability metric with point and interval hypothesis testing under both the frequentist and Bayesian perspectives. They found that adopting the hypothesis testing approach may lead to some inconsistencies as to whether

the model is deemed valid or not, whereas the model reliability metric dos not present these inconsistencies. Li and Mahadevan [91] expand this model reliability metric to take into account multivariate outputs and Mullins et al. [32] further investigate the use of the model reliability metric for validation. To compute the probability γ, one needs to know the distribution of the discrepancy E. We assume here that the random variables $Y_{\text{exp}}(x_{\text{cal}}, z)$ and Θ are independent. This assumption is reasonable since the first quantity relates to the experimental observations while the second quantity is related to the various hypotheses on the model. We can then estimate $\gamma = F_{|E|}(\varepsilon)$ with any of the previously mentioned sampling methods by performing the sampling of the empirical distributions of $Y_{\text{exp}}(x_{\text{cal}}, z)$ and of the model parameters Θ independently.

Validation Scenario

The choice of the validation scenario x_{val}, the type of observable, and the value of the sensor parameters z_{val} have a significant impact on the actual outcome of the validation process, as it will be illustrated in Section 3.4.1. It is seldom the case that we can replicate in a controlled environment the conditions under which we would like to perform the prediction, wherein the prediction and validation scenario are the same. If it is not the case, then we must come up with an *informative* and *relevant* validation scenario with respect to the prediction scenario. Moreover, if the QoI is not observable, then we need to determine which type of observables and which sensor parameters z should be employed for the validation. These are the main issues addressed in this research work, which will be further explored in Section 3.3.

Validation Workflow

A complementary objective of our work is to eliminate as much as possible arbitrary decisions in the validation process. Indeed, by providing a way to determine the validation scenario x_{val}, the type of observables, and the sensor parameters z_{val}, we are closer to *automating* the validation process. Our perspective of what a validation workflow should look like is summarized by Algorithm 1.

The first step consists in designing an appropriate validation scenario. One would then carry out the validation experiment and estimate the validation metric to verify that its prediction indeed reflects the reality. As a final remark, we do not address in this work the propagation of the modeling errors to the QoI at the prediction scenario, nor the decision making about the model validity. Our objective is mainly to search for the best validation experiments so that one can better assess the capability of the model to predict the QoI. That said, we conjecture that our approach for designing optimal validation experiments, as described in

Algorithm 1: Validation Process

Input: Model (r, p, u), quantity of interest h_{qoi}, prediction scenario x_{pred}, choice of observation functional h_{obs}, choice of the error $E(x, z)$, model parameters Θ, constrained sets $\mathcal{X}_{\text{lab}}$ and $\mathcal{Z}_{\text{lab}}$, error tolerance ε, threshold η

Computation of the validation scenario x_{val}, sensor parameters z_{val}, and type of observation functional $h_{\text{obs,val}}$ (Algorithm 2);
Realization of the experiment to obtain $y_{\text{exp}}(x_{\text{val}}, z_{\text{val}})$;
Computation of the validation metric (either $d(F_Y, F_{Y_{\text{exp}}})$ or γ);
if $d(F_Y, F_{Y_{exp}}) \leq d_{tol}$ *or* $\gamma \geq \eta$ **then**
    Model is not invalidated
else
    Model is invalidated
end

the next section, complements the validation workflow presented in Algorithm 1 to provide useful assessments of the model validity.

3.3 Optimal Design of Validation Experiments

A validation scenario to obtain validation data must be *representative* of the prediction scenario, since the objective is to assert whether or not the model is valid for predictive purposes, more specifically, for the prediction of the quantity of interest $Q = h_{\text{qoi}}(x_{\text{pred}}, \Theta))$. By *representative*, we mean that the various hypotheses on the model should be similarly satisfied in both the prediction and validation scenarios. Our view is that the behavior of the model with respect to the parameters and their uncertainty should be as similar as possible in the validation and prediction scenarios. For example, if a parameter significantly influences the QoI at the prediction scenario, then the observation functional at the validation scenario should also be significantly influenced by the same parameter. Vice versa, if a parameter does not influence the QoI at the prediction scenario, then it must also be the case for the observation functional at the validation scenario. In other words, an analysis of the parameter sensitivity should guide the optimal design of validation experiments.

3.3.1 Sensitivity Analysis

Sensitivity analysis is a vast subject with multiple objectives and methodologies, see e.g. [50, 52, 95] for thorough overviews on the topic. Examples of problems addressed by sensitivity analysis are:

- Characterization of the relationship between the parameters and the model outputs.

- Identification of non-influential parameters, a topic often referred in the literature to as *factor fixing*.
- Identification of the parameters that influence the most the model outputs, usually referred to as *factor prioritization*. One objective here is to further reduce the uncertainty affecting these parameters.

Sensitivity methods are usually classified as either local or global. On one hand, local methods compute the sensitivity (e.g. derivatives) around a specific point in the parameter space and do not take into account the uncertainty affecting the parameters. On the other hand, global methods incorporate the uncertainty in the parameters. The Sobol'-based sensitivity analysis is a global approach that computes the so-called Sobol' sensitivity indices, which quantitifies how much of the variance in a model output can be attributed by a single parameter or a subset of parameters [57].

We consider in our approach that the model parameters Θ are uncertain so that we need to employ a global sensitivity analysis method. Moreover, our goal being to assess the influence of the various parameters on the model outputs, either the observables or the QoI (the first objective aforementioned), we seek a description of the response surface of the observation and QoI functionals with respect to the parameters. Our objective is thus different from that of allocating the variance of $Y = h_{\text{obs}}(x, \Theta, z)$ or $Q = h_{\text{qoi}}(x, \Theta)$ to parameters (or subsets of parameters) such as in the Sobol' approach.

3.3.2 Active Subspace Method

One sensitivity method that allows the characterization of a response surface while taking into account the uncertainty of the parameters is the Active Subspace method [65, 66]. The method computes the gradient of a model functional h with respect to the control and model parameters to quantify their influence. The expectation of the outer product of the gradient with respect to the distribution F_Θ is performed to assemble the so-called *influence matrix* M_h

$$\begin{aligned} M_h(x,z) &= \mathbb{E}\left(\nabla_{(x,\theta)} h(x,\Theta,z)\, \nabla_{(x,\theta)} h(x,\Theta,z)^T\right) \\ &= \text{Cov}\left(\nabla_{(x,\theta)} h(x,\Theta,z)\right) + \mathbb{E}\left(\nabla_{(x,\theta)} h(x,\Theta,z)\right) \mathbb{E}\left(\nabla_{(x,\theta)} h(x,\Theta,z)\right)^T. \end{aligned} \tag{3.11}$$

The influence matrix $M_h(x,z)$ belongs to $\mathcal{S}_+^d$, the set of positive semi-definite matrices of size d. One proceeds with the eigenvalue decomposition of $M_h(x,z)$ to identify the principle directions influencing the outputs of the model. This influence matrix is characterized by large variations in the gradient and/or large expected gradients. From now on, the influence matrix M_h will constitute our quantitative description of the surface response of h.

Given two influence matrices M_1 and M_2 characterized by different functionals h, different scenarios x, and different sensor parameters z, we define the distance between the two influence matrices as

$$d(M_1, M_2) = \|M_1 - M_2\|_2, \tag{3.12}$$

where $\|M\|_2$ denotes the spectral norm of M (other norms could have been considered as well). The distance essentially measures the error between an influence matrix M_1 and a target influence matrix M_2. We note that the use of the Active Subspace method presumes enough regularity of the observations and QoI functionals.

Having defined a quantification of the response surface of a functional as well as a means for comparison, we turn our attention to the problem at hand, that of defining a validation experiment specifically tailored toward the prediction of the QoI at a given prediction scenario.

3.3.3 Optimal Choice of the Validation Scenario

We describe the set of controlled experiments as the set of scenarios $\mathcal{X}_{\text{lab}} \subseteq \mathcal{X}$ for which one is able to make a laboratory experiment to interrogate the system of interest. We recall that the prediction scenario x_{pred} may or may not belong to $\mathcal{X}_{\text{lab}}$. We thus formulate two requirements regarding an optimal validation scenario:

1. If $x_{\text{pred}} \in \mathcal{X}_{\text{lab}}$, then the validation scenario should be given by $x_{\text{val}} = x_{\text{pred}}$.
2. If the QoI functional h_{qoi} is in fact an observable h_{obs}, then the observable and the sensor parameters employed for the validation should correspond to those of the QoI.

These two requirements will guide the definition of the two optimal design problems that will be set up to specify the goal-oriented validation experiment. The first step in specifying the validation experiment consists in determining the validation scenario x_{val}. We recall here that the QoI functional h_{qoi} is only a function of the control parameters x and of the model parameters Θ. For the sake of simplicity, we consider that the model parameters are identically distributed between the prediction and validation scenarios. Hence, we have two random quantities of interest, $Q_{\text{val}} = h_{\text{qoi}}(x_{\text{val}}, \Theta)$ and $Q_{\text{pred}} = h_{\text{qoi}}(x_{\text{pred}}, \Theta)$. However, one could easily alleviate this assumption by considering different distributions for the model parameters at the prediction and validation scenarios.

We seek the validation scenario x_{val} that minimizes the distance (3.12), with the constraint

that the scenario be realizable in a controlled environment

$$x_{\text{val}} \in \underset{x \in \mathcal{X}_{\text{lab}}}{\operatorname{argmin}}\, d(M_{h_{\text{qoi}}}(x), M_{h_{\text{qoi}}}(x_{\text{pred}})). \tag{3.13}$$

In other words, we look for the validation scenario such that the response surface of Q_{val} (encoded by the influence matrix $M_{h_{\text{qoi}}}(x_{\text{val}})$) resembles the most the response surface of Q_{pred} for the prediction scenario. This is another way of saying that the various model hypotheses and simplifications influence in a similar fashion the QoI in the validation and predictive settings. We observe that if $x_{\text{pred}} \in \mathcal{X}_{\text{lab}}$, then a global minimum with zero value is attained at $x_{\text{pred}} = x_{\text{val}}$, since the objective function in (3.13) is non-negative. However, the global minimum may not be unique. Indeed, a linear functional h_{qoi} with respect to the parameters features the same influence matrix $M_{h_{\text{qoi}}}(x)$ whatever the value of these parameters. Hence, the first requirement stated above is partially fulfilled, since the optimization problem (3.13) may not possess a unique minimum.

3.3.4 Optimal Choice of Observables and Sensor Parameters

The goal in this section is to define the problem whose solution will provide an observation functional h_{obs} and sensor parameters z in order to fully characterize the optimal validation experiment. We recall that QoIs are not necessarily observable either in the prediction or the validation scenarios. We nevertheless want to validate the model for the prediction of the QoI. With an optimal prediction scenario obtained by solving (3.13), we now need to determine the functional h_{obs} and sensor parameters z that best mimic the influence matrix of $Q_{\text{val}} = h_{\text{qoi}}(x_{\text{val}}, \Theta)$. Again, the rationale is that the influence matrix reflects the impact of the various hypotheses and assumptions of the model. The result of the validation process obtained with this particular functional h_{obs} and sensor parameters z could then be reasonably extrapolated to the QoI. Since the choices of the observation functional and of the sensor parameters z are constrained by the available experimental capabilities, we introduce here the set $\mathcal{H}_{\text{lab}}$ of possible observation functionals and the set $\mathcal{Z}_{\text{lab}}$ of possible sensor parameters. The optimal design problem to be solved is then given by

$$(h_{\text{obs,val}}, z_{\text{val}}) \in \underset{(h_{\text{obs}}, z) \in \mathcal{H}_{\text{lab}} \times \mathcal{Z}_{\text{lab}}}{\operatorname{argmin}} d\left(\frac{M_{h_{\text{obs}}}(x_{\text{val}}, z)}{\operatorname{Tr}(M_{h_{\text{obs}}}(x_{\text{val}}, z))}, \frac{M_{h_{\text{qoi}}}(x_{\text{val}})}{\operatorname{Tr}(M_{h_{\text{qoi}}}(x_{\text{val}}))} \right). \tag{3.14}$$

The normalization of the influence matrices for the observational and QoI functional is required since both functionals may represent different physical quantities with different scaling and units. The validation experiment tailored toward the prediction of the QoI is now fully specified. It consists in interrogating the system of interest for the experimental observa-

tion corresponding to the observation functional $h_{\text{obs,val}}$ at the sensor parameters z_{val} under the scenario x_{val}. The whole process to search for the optimal experiment is described in Algorithm 2.

The values of the objective functionals of Problems (3.13) and (3.14) both provide useful information regarding the quality of the validation experiment. If both objective functionals are small, then the influence matrix of the observables h_{obs} at the validation scenario mimics well the influence matrix of the QoI at the prediction scenario. Since the influence matrices provide a means to quantify the effect of the various hypotheses made about the model on the prediction, one should now be more confident that the observable at the validation scenario expresses the same hypotheses as the QoI at the prediction scenario. It may happen that the modeling errors in the QoI at the prediction scenario are different from the modeling errors in the observable at the validation scenario. However, we ensure in this way that the differences mainly arise from other physical phenomena that are not captured by the model, rather than from differences between the behavior of the observation at the validation scenario and that of the QoI at the prediction scenario. On the other hand, if the objective functional of problem (3.13) is large, then one may need to consider expanding the set of controlled scenarios $\mathcal{X}_{\text{lab}}$. Also, if the objective functional of Problem (3.14) is large, then one may seek other types of experimental measurements. In case of one or both situations, one must exercise caution in extrapolating the result of the validation process to the prediction setting.

The proposed approach to design optimal validation experiments differs from those previously described to design optimal calibration experiments. It is important to acknowledge that their respective objectives are quite different. The objective here is to design an optimal validation scenario such that the behavior of the observable resembles the most the behavior of the QoI at the prediction scenario. When designing optimal calibration experiments, the main goal is to find scenarios such that the model parameters Θ can be better estimated. In other words, the proposed approach for the design of optimal validation experiments is not intended at updating the knowledge of the model parameters Θ.

The solution of the optimization problems (3.13) and (3.14) is performed with the solver NOMAD [74] and its use is further detailed in 3.C.

3.4 Numerical Examples

We illustrate our methodology on two examples, the first one consisting in the projectile problem introduced in Section 3.2.1 and the second one consisting in a pollutant transport problem. The first problem is used to illustrate the entire validation process and to emphasize

Algorithm 2: Optimal Design of Validation Experiment

Input: Model (r, p, u), quantity of interest h_{qoi}, prediction scenario x_{pred}, model parameter Θ, set of controlled scenarios $\mathcal{X}_{\text{lab}}$, set of observation functionals $\mathcal{H}_{\text{lab}}$, set of controlled sensor parameters $\mathcal{Z}_{\text{lab}}$

1 Computation of the validation scenario x_{val} by solving (3.13);

2 Computation of the validation observation functional $h_{\text{obs,val}}$ and validation sensor parameter z_{val} by solving (3.14);

Output: Validation scenario x_{val}, validation observation functional $h_{\text{obs,val}}$, validation sensor parameters z_{val}

the importance of the design of the validation experiment. In particular, we will provide an example of a poorly chosen validation experiment that leads to a false positive, that is, the model will not be deemed invalid when it should be at the given prediction scenario. By contrast, we will show that the application of our methodology to design an optimal validation experiment produces a true negative. The second problem will show the importance of the definition of the QoI on the validation experiment.

3.4.1 The Projectile Problem

The first example considers the system of interest described in Section 3.2.1, where we seek the maximum altitude reached by a spherical projectile when launched vertically. The model is given by the linear ordinary differential equation (3.2) for which one can analytically compute the value of the QoI as

$$q(x, \theta) = \max_t u(t) = u_0 + \frac{m v_0}{3\pi \exp(\mu)\ell} - \frac{m^2 g}{(3\pi \exp(\mu)\ell)^2} \ln\left(1 + \frac{3\pi \exp(\mu)\ell v_0}{mg}\right). \tag{3.15}$$

We recall that the control parameters are given here by $x = (m, \ell, u_0, v_0)$ while the model parameters are $\theta = (g, \mu)$. Since the proposed QoI is non-linear with respect to some of the control parameters, the associated influence matrix (3.11) depends on the scenario x. Moreover, we set the model parameters to $\Theta = (G, U) \sim (\mathrm{N}(9.81, 0.3^2), \mathrm{N}(-5\ln(10), 0.5^2))$ independently, where N denotes normal distributions.

Optimal Validation Scenario

To illustrate the possibility that a poorly designed validation experiment may yield a false positive, we need to consider a specific prediction scenario x_{pred} and a specific controlled environment $\mathcal{X}_{\text{lab}}$ that describes the possible validation scenarios x_{val}. The prediction scenario, the tentative validation scenario, and the controlled environment are described in Table 3.2.

Table 3.2 Prediction scenario x_{pred}, controlled environment $\mathcal{X}_{\text{lab}}$, and tentative validation scenario $x_{\text{val,tent}}$

	Mass m	**Diameter** ℓ	**Initial Position** u_0	**Initial Velocity** v_0
Prediction Scenario x_{pred}	0.05	0.01	1	100
Controlled Environment $\mathcal{X}_{\text{lab}}$	[1,5]	[0.01,0.1]	[0,2]	[10,105]
Tentative Validation Scenario $x_{\text{val,tent}}$	3	0.055	1	57.5

The controlled environment $\mathcal{X}_{\text{lab}}$ consists in the Cartesian product of the intervals for the parameters x, as defined in Table 3.2. The tentative validation scenario is chosen to be at the center of the controlled environment set $\mathcal{X}_{\text{lab}}$, thus experimentally producible. This tentative choice could have been more informed, with expert opinion for example, but we will consider this tentative validation scenario for illustration purpose.

We now solve problem (3.13) to obtain the optimal validation scenario for this predictive setting and obtain $x_{\text{val}} = (1.000, 0.100, 0.705, 100.446)$. It is interesting and important to observe that this optimal validation scenario is not the closest point of $\mathcal{X}_{\text{lab}}$ (in the l_2 norm) to the prediction scenario. Figure 3.3 compares the influence matrices M_h (3.11) of the QoI (3.15) for the prediction scenario x_{pred}, the tentative validation scenario $x_{\text{val,tent}}$, and the aforementioned optimal validation scenario x_{val}.

We observe that the u_0 and v_0 components of the first eigenvector v_1 have slightly changed between $x_{\text{val,tent}}$ and x_{val} in order to resemble more closely the one of the prediction scenario. Its associated eigenvalue has also slightly increased to better fit the one at the prediction scenario. The second and third eigenvectors seem inverted between the prediction and tentative validation scenario. The importance of this inversion is however reduced at the optimal validation scenario since the second and third eigenvalues are almost identical. Moreover, we observe that the second eigenvector of the optimal validation scenario matches more closely the third eigenvector of the prediction scenario. Also, the influence of the mass m seems better captured in the optimal validation scenario than in the tentative validation scenario. All things considered, the influence matrix of the QoI at x_{val} better mimics the influence matrix of the QoI at x_{pred} than the tentative validation scenario.

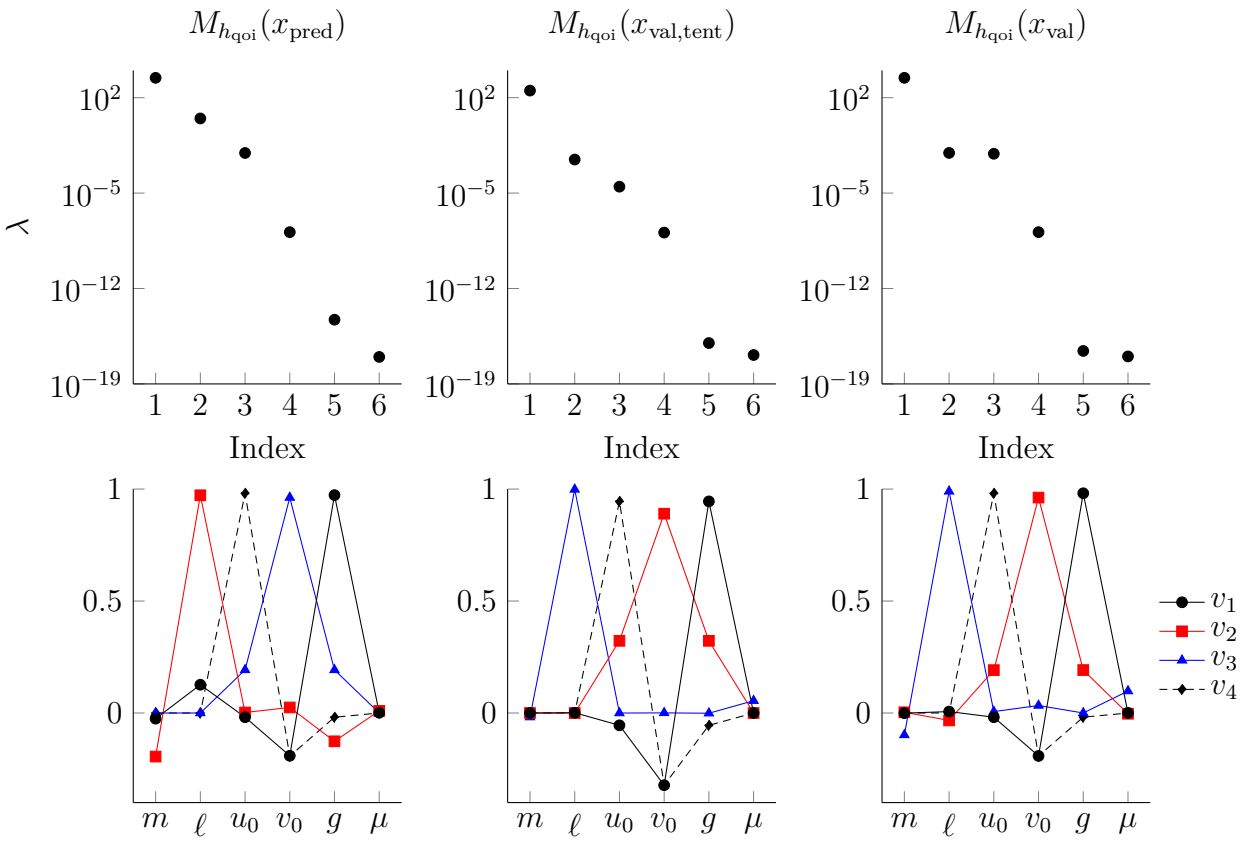


Figure 3.3 Eigenvalues (top row) and eigenvectors (bottom row) associated with the influence matrices of the QoI (3.15) for the prediction scenario x_{pred} (left), the tentative validation scenario $x_{\text{val,tent}}$ (center), and the optimal validation scenario x_{val} (right). For the sake of clarity, only the first four eigenvectors associated with the largest four eigenvalues are displayed as the values of the other two eigenvalues are negligible.

Optimal Sensors and Observation Functional

We search for the best observation functional h_{obs} between the functionals $h_{\text{obs}}(x_{\text{val}}, \Theta, z) := u$ and $h_{\text{obs}}(x_{\text{val}}, \Theta, z) := \mathrm{d}^2u/\mathrm{d}t^2$, as functions of time t, and the best sensor parameters $z := t$ that should be employed in the validation experiment. In order to do so, we evaluate the objective functional of Problem (3.14) for both observation functionals and for various times t under the validation scenario x_{val} computed previously. The results are shown in Figure 3.4.

As expected, there is a specific time t for which the influence matrix of $h_{\text{obs}}(x_{\text{val}}, \Theta, z) := u(t)$ is exactly the same as the influence matrix of $h_{\text{qoi}}(x_{\text{val}}, \Theta)$. This particular time t corresponds to the instant at which the projectile has reached the maximal altitude (as shown in Figure 3.5). It is interesting to observe in Figure 3.4 that the objective functionals for the position u and acceleration $\mathrm{d}^2u/\mathrm{d}t^2$ evolve quite differently in time. On one hand,

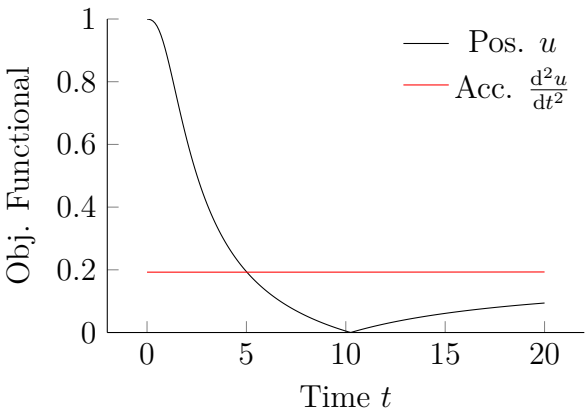


Figure 3.4 Objective functional of Problem (3.14) associated with the observation functionals $h_{\text{obs}}(x_{\text{val}}, \Theta, z) \coloneqq u$ and $h_{\text{obs}}(x_{\text{val}}, \Theta, z) \coloneqq \mathrm{d}^2 u / \mathrm{d}t^2$.

the influence matrix for the acceleration remains quite different from the influence matrix of the QoI at the validation scenario for all times t. On the other hand, the objective functional (3.14) for the position u at small times t is quite large. This is explained by the fact that the position right after the launch is mainly influenced by the initial position u_0, which is not the case for the QoI at the validation scenario, see rightmost plot on bottom row in Figure 3.3. However, a rapid decrease in the objective functional of Problem (3.14) for the position indicates that the influence of the initial position (alongside the other parameters) quickly tend to that of the QoI.

For the sake of presenting the whole validation process, and to illustrate the possibility of obtaining a false validation result, we need some experimental observations. We shall use manufactured or synthetic data provided by what we will refer here to as a *fine model* that serves as a surrogate to the physical system of interest. Indeed, the precise source of experimental observations is not directly relevant here, since they do not play any role in the design of the optimal validation experiment.

Manufactured Data

The fine model is provided by the following non-linear ODE [94, Section 7.6] and initial conditions

$$m\frac{\mathrm{d}^2\tilde{u}}{\mathrm{d}t^2}(t) + \frac{\rho\pi\ell^2 c_D}{8}\frac{\mathrm{d}\tilde{u}}{\mathrm{d}t}(t)\left|\frac{\mathrm{d}\tilde{u}}{\mathrm{d}t}(t)\right| = -mg, \qquad \text{for } t \in (0,T), \tag{3.16a}$$

$$\tilde{u}(0) = u_0, \tag{3.16b}$$

$$\frac{\mathrm{d}\tilde{u}}{\mathrm{d}t}(0) = v_0. \tag{3.16c}$$

Following [104], the definition of the friction coefficient c_D depends on the Reynolds number Re

$$\mathrm{Re}(t) = \frac{\rho\ell}{\exp(\mu)}\left|\frac{\mathrm{d}u}{\mathrm{d}t}(t)\right|, \tag{3.17}$$

where ρ is the density of the ambient air. The friction coefficient c_D is defined for laminar regimes, i.e., $\mathrm{Re} < 2 \times 10^5$, as:

$$c_D = \frac{24}{\mathrm{Re}}\left(1 + 0.15 \times \mathrm{Re}^{0.681}\right) + \frac{0.407}{1 + 8710 \times \mathrm{Re}^{-1}}. \tag{3.18}$$

In case of very low Reynolds numbers ($\mathrm{Re} \ll 1$), the coefficient is well approximated by $c_D = 24/\mathrm{Re}$. With this approximation, one actually recovers the linear model (3.2).

The fine model allows us to obtain measurements of the position $\tilde{u}$ at any time t. Realizations $y_{\mathrm{exp}}(x,z)$ of these pseudo-experimental observations $Y_{\mathrm{exp}}(x,z)$ are obtained by sampling

$$Y_{\mathrm{exp}}(x,z) \sim \tilde{u}(x,z)\,\mathrm{N}(1, 0.05^2). \tag{3.19}$$

This choice for the noise is arbitrary and corresponds to a multiplicative Gaussian noise with standard deviation 5%. We suppose here that the discretization errors are precisely controlled while solving the ODE (3.16). Finally, for this fine model, we fix the values of the gravitational constant, the air viscosity, and the air density to $g = 9.81$, $\exp(\mu) = 1.8 \times 10^{-5}$, and $\rho = 1.2$, respectively.

Validation Analysis

We now compare the trajectories obtained with Model (3.2) and Model (3.16) for the prediction scenario x_{pred}, the tentative validation scenario $x_{\mathrm{val,tent}}$, and the optimal validation scenario x_{val}. We adopt here the model reliability metric to perform the validation. Hence, we compute the discrepancy (3.5) for the three scenarios at their respective optimal sensor

parameter z_{val}, which corresponds to the time the projectile reaches the maximal altitude. The trajectories and distributions $F_{|E|}$ of the discrepancy E are shown in Figure 3.5. It is important to mention that the distribution of the discrepancy E and the trajectory of the physical system (red curve) cannot in general be accessed for the prediction scenario since experimental data are not available in that case. We can produce those plots here only because the measurements are obtained with the help of the model (3.16).

We observe that our model (3.2) poorly captures the maximum altitude reached by the projectile at the prediction scenario, since there is a significant difference between the red and blue curves, as shown in the top left plot of Figure 3.5. The friction forces in the model are in fact underestimated. This mismatch between the model and the system of interest is reflected in the distribution $F_{|E|}$ of the discrepancy: the probability that the absolute error on the maximal position be less than 80 is close to zero. In other words, one is almost sure that the absolute error on the maximal position will be between 80 and 110. The model for the prediction of the maximal altitude under the prediction scenario x_{pred} can thus be considered invalid. Although this information is not available in practice, the validation process should inform us beforehand on the accuracy of the model predictions.

Under the tentative validation scenario $x_{\text{val,tent}}$, we observe a relatively good agreement between the position of the projectile for the model and for the system of interest. This can be explained by the fact that the friction force is less important with respect to the inertial and gravitational forces. Since the only difference between the model and the system of interest is due to the modeling of the friction force, the discrepancy E, computed at the optimal sensor point $z = t \approx 5.8$, remains small. Its distribution $F_{|E|}$ indicates that the observational error is almost surely less than 10. Depending on the tolerance ε and threshold η on the validation metric, the model would likely have been deemed valid. Hence, the model is not appropriate for the prediction of the QoI at the prediction scenario and the validation process using this tentative validation scenario would have given a false positive.

In the case of the optimal validation scenario x_{val}, we observe a significant departure between the model and the system of interest. The optimal validation scenario seems to capture the fact that the friction forces in our model are underestimated. With the optimal sensor point previously computed, the discrepancy indicates that the error on the position is almost surely between 280 and 310. The model would likely have failed the validation test, which better reflects the real predictive capabilities of the model under the prediction scenario.

We would like to emphasize here that the proposed methodology to design an optimal validation experiment does not guarantee that the outcome of the validation process will always provide one with a correct answer. It only ascertains that the influence matrix of the obser-

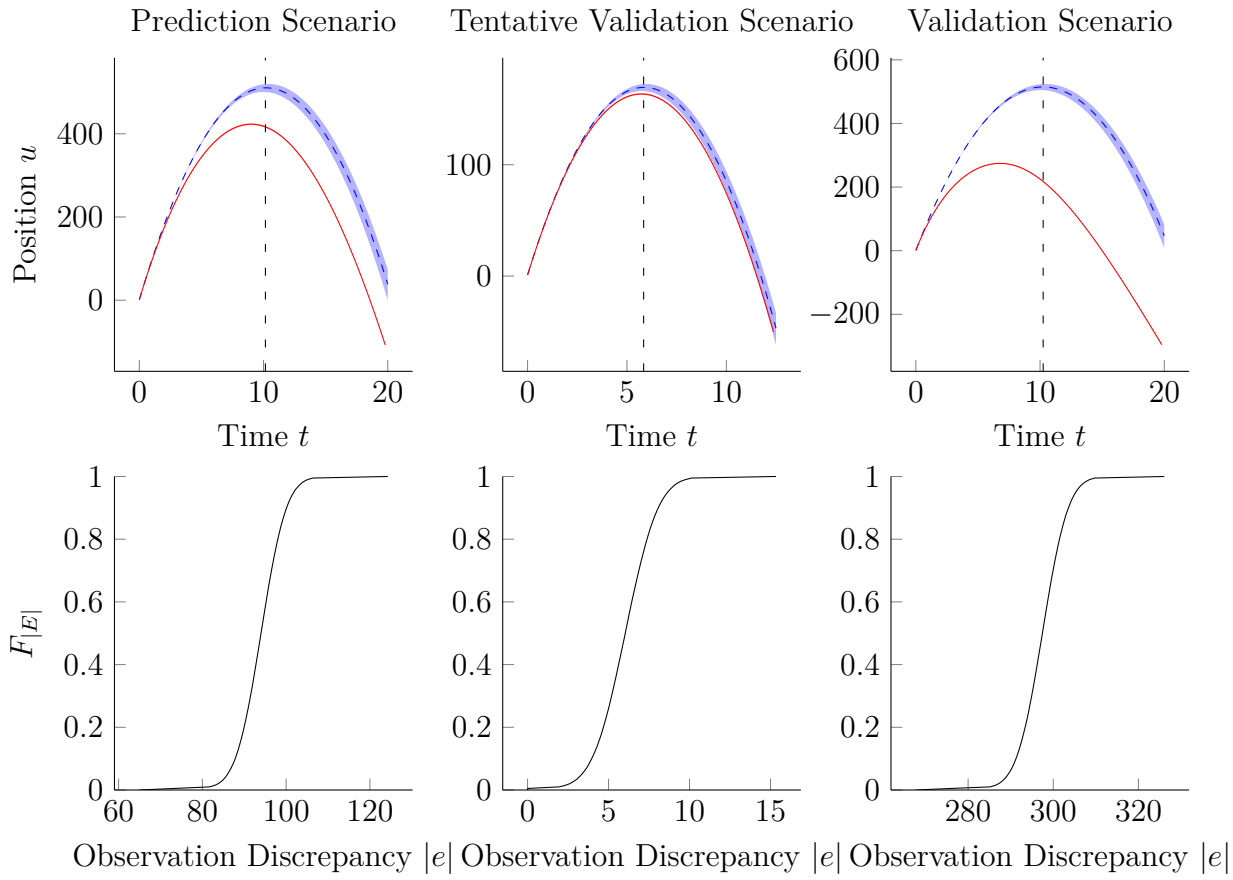


Figure 3.5 Position u (top row) and discrepancies (bottom row) for the prediction scenario (left column), tentative validation scenario (center column), and the optimal validation scenario (right column). The experimental observations are in red and the model observations are in blue. The shaded areas represent the 95% confidence interval for both the experimental and model observations. The absolute value of the discrepancies (3.5) are computed at the sensor location indicated by the black dashed lines.

vation at the validation scenario is as much as possible comparable to the influence matrix of the QoI at the prediction scenario. In this example, it happens that the conditions of the optimal validation experiment allow the model to capture the relative importance of the friction forces with respect to the other forces, as in the prediction scenario. In other words, since the friction forces are not correctly modeled in the validation setting, one may conclude that it could also be the case for the prediction scenario, which is indeed true here.

3.4.2 Contaminant Transport Problem

The second example consists in a problem of pollutant transport in a fictitious river where we wish to analyze the impact of the design of a new factory upstream. We consider two QoIs,

both representing the mean concentration of contaminant in a specific region of the domain. The model we wish to validate is a 2D linear steady-state diffusion-advection equation that governs the concentration $\phi = \phi(z)$ of the contaminant in the river:

$$-\nabla \cdot (\exp(k)\nabla\phi) + \nabla \cdot (v\phi) = 0, \qquad \text{in } \Omega, \tag{3.20a}$$

$$\phi = \phi_D, \qquad \text{on } \Gamma_{\text{west}}, \tag{3.20b}$$

$$-n \cdot \exp(k)\nabla\phi = 0, \qquad \text{on } \partial\Omega \setminus \Gamma_{\text{west}}, \tag{3.20c}$$

where n denotes the outward unit normal to the domain boundary, $\exp(k)$ is the diffusivity coefficient, and v is the advection velocity. For the sake of simplicity, we consider the velocity field v to be given and known, so that it is not considered a parameter in this study (see 3.A).

The control parameters x consist in the parameters that characterize the Dirichlet condition ϕ_D, to be described later, and the model parameter is only $\theta = k$. We suppose here that the diffusivity parameter k is well-known so that $\Theta \sim \delta(k_0)$, with δ the Dirac measure and $k_0 = -2\ln(10)$. The sensor parameters are the Cartesian coordinates $z = (z_1, z_2)$. A sketch of the geometry is provided in Figure 3.6. The domain consists in an inlet and an outlet on the west and east boundaries, respectively. Two docks hinder the flow around the first region of interest, denoted by Ω_1. The second region of interest Ω_2 is shown in Figure 3.6. We note that the particular dimensions of the domain Ω and the particular value of the model parameter k are arbitrarily chosen, as the example is for illustrative purpose only.

To validate the model, we imagine one can perform an experiment consisting in injecting a small and controlled amount of contaminant upstream. The Dirichlet condition ϕ_D is therefore used to describe both the pollutant release from the factory and the validation experiment. We parametrize ϕ_D in terms of the mollifier centered at $z_2 = z_0$, of length L and intensity c, as follows:

$$\phi_D(z_0, L, c, z_2) = \begin{cases} c\exp\left(\left(\left|\frac{z_2 - z_0}{L}\right| - 1\right)^{-1}\right), & \text{if } \left|\frac{z_2 - z_0}{L}\right| < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{3.21}$$

The control parameters in this application are given by $x = (z_0, L, c)$. This parametrization is sufficiently rich to describe both the pollutant release from the factory and a wide range of validation scenarios, while involving a limited number of parameters. Moreover, the use of above mollifier is attractive thanks to its regularity properties, since we need to compute the gradient of functionals with respect to the control parameter x (as explained in 3.B).

We can now specify both the prediction scenario x_{pred} and the controlled environment $\mathcal{X}_{\text{lab}}$.

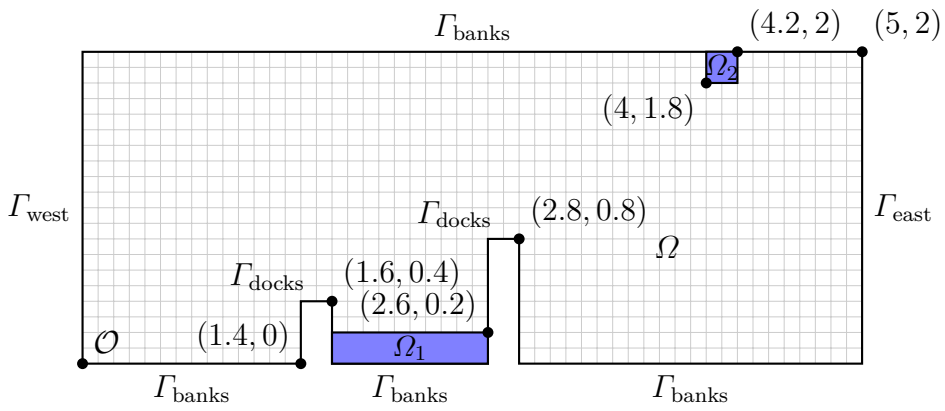


Figure 3.6 Domain Ω and its boundaries for the contaminant transport model. The regions for which we seek the mean pollutant concentration are indicated by Ω_1 and Ω_2 (respectively employed for the definition of the QoIs). The grid is used to indicate the regions Ω_{obs} in which the mean concentration of pollutant can be observed.

The prediction scenario consists in a description of the possible pollutant profile from the factory to be built upstream and is provided in Table 3.3. The controlled environment $\mathcal{X}_{\text{lab}}$ includes all validation scenarios that can actually be performed. We suppose here that the pollutant can be injected in a controlled manner anywhere along the boundary Γ_{west}, while the length L and intensity c can only take limited values. The intervals for the three control parameters z_0, L, and c are also provided in Table 3.3.

We also consider the two QoIs given by

$$h_{\text{qoi},i}(x, \theta) = \frac{1}{|\Omega_i|} \int_{\Omega_i} \phi(x, \theta, z) \, \mathrm{d}z, \qquad \text{for } i = 1, 2,$$

where Ω_i, $i = 1, 2$, are the two sub-regions shown in Figure 3.6.

Optimal Validation Scenarios

Now that the prediction scenario x_{pred}, quantities of interest $h_{\text{qoi},i}$, and controlled environment $\mathcal{X}_{\text{lab}}$ are prescribed, we can compute the optimal validation scenarios with respect to the two QoIs. We thus solve Problem (3.13) for each QoI. The concentration of the pollutant for the prediction scenario x_{pred} and the optimal validation scenarios with respect to the first and second QoI are all shown in Figure 3.7.

We observe that the optimal validation scenarios are highly sensitive to the particular QoI we wish to predict since the location z_0 of the pollutant varies considerably between the two QoIs. This result is important as it highlights the fact that it is the combination of

Table 3.3 Description of the prediction scenario x_{pred} and of the controlled environment $\mathcal{X}_{\text{lab}}$ for the pollutant transport model.

	Position z_0	**Length** L	**Intensity** c
Prediction Scenario x_{pred}	0.75	0.6	2
Controlled Environment $\mathcal{X}_{\text{lab}}$	[0.1,1.9]	[0.05,0.1]	[0.5,1]

the prediction scenario and the QoI one wants to predict that should dictate the validation scenario. From the prediction scenario, we observe that the pollutant enters the domain Ω_1 associated with the first QoI, while very little pollutant reaches the second domain Ω_2 for the second QoI. We reiterate here that the objective is not to find the validation scenario x_{val} such that $h_{\text{qoi}}(x_{\text{val}}, \Theta) \approx h_{\text{qoi}}(x_{\text{pred}}, \Theta)$, but the scenario such that $M_{h_{\text{qoi}}}(x_{\text{val}}) \approx M_{h_{\text{qoi}}}(x_{\text{pred}})$.

The influence matrices $M_{h_{\text{qoi}}}(x_{\text{val}})$ and $M_{h_{\text{qoi}}}(x_{\text{pred}})$ are characterized by their eigenvalues and eigenvectors shown in Figure 3.8. The absence of uncertainty in the model parameter $\theta = k$ implies that the influence matrices $M_{h_{\text{qoi}}}$ should be of rank one, since the covariance in Equation (3.11) is null in this case. This is clearly reflected in the numerical results, where the first eigenvalue of each influence matrix are several orders of magnitude higher than the others.

Concerning the first QoI, we observe that the magnitudes of the eigenvalues are slightly smaller for the validation than those for the prediction. This can be explained by the fact that the overall magnitude of the pollutant is smaller in the validation setting because of the constraint on the controlled environment $\mathcal{X}_{\text{lab}}$. We also observe that the eigenvector for the optimal validation scenario tends to match as closely as possible the eigenvector of the prediction scenario. One can explain this difference by remarking that the size L of the region in which the pollutant is injected should have more influence in the validation scenario since the quantity of injected pollutant is smaller than for the prediction scenario. The optimal validation scenario actually succeeds in capturing the fact that the intensity c of the source of pollutant and the diffusivity k are not very influential in determining the mean pollutant concentration in Ω_1.

Concerning the second QoI, we observe a very good agreement between the eigenvalues of the influence matrix at the validation scenario and those at the prediction scenario. As with the first QoI, the eigenvector for the validation scenario tends to match as closely as possible the eigenvector for the prediction scenario.

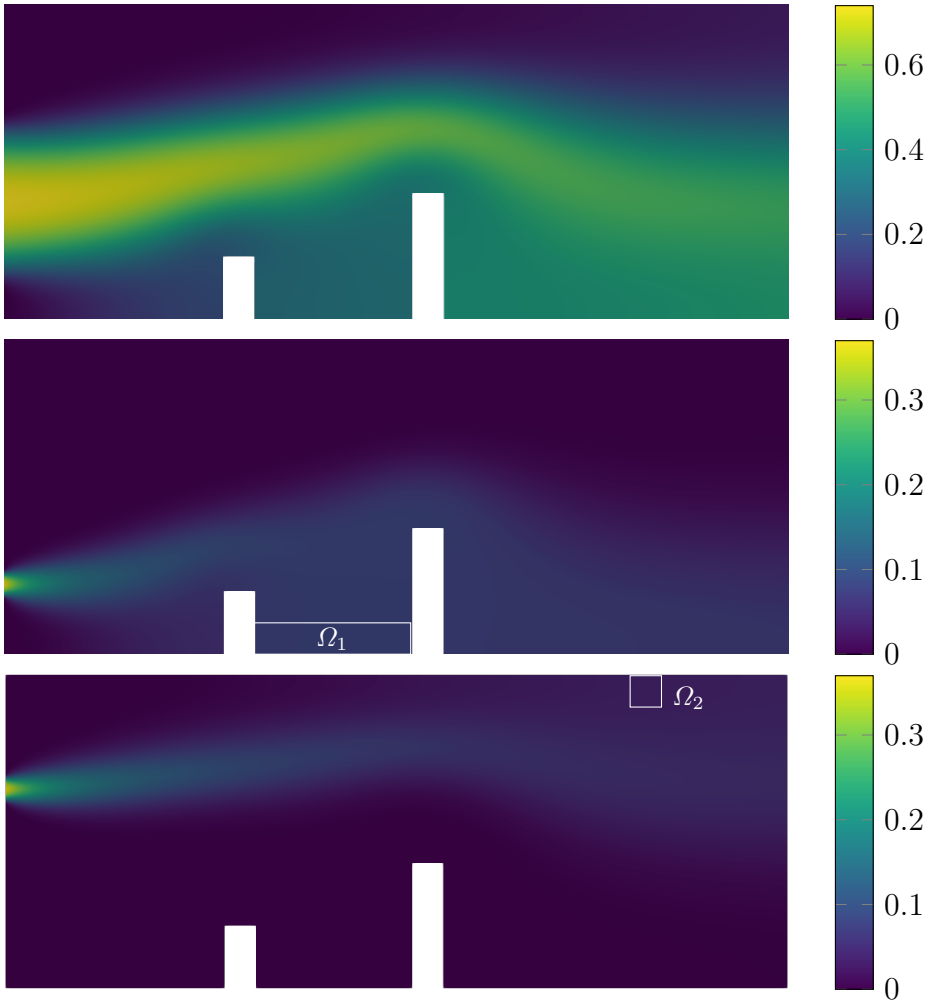


Figure 3.7 Concentration of the pollutant for the prediction and optimal validation scenarios: (top) prediction scenario, (center) optimal validation scenario associated with $h_{\text{qoi},1}$, (bottom) optimal validation scenario associated with $h_{\text{qoi},2}$.

Positioning of the Sensors

Once the optimal validation scenario has been computed for each quantity of interest considered, we tackle the last step of our methodology, namely the optimal positioning of a sensor. For the pollutant transport problem, we consider only one type of observation functional h_{obs}

$$h_{\text{obs}}(x, \theta, z) = \frac{1}{|\Omega_{\text{obs}}(z)|} \int_{\Omega_{\text{obs}}(z)} \phi(x, \theta, w) \, \mathrm{d}w, \tag{3.22}$$

where $\Omega_{\text{obs}}(z)$ represents a square region centered at z of length and width equal to 0.1. The objective is now to find the location z of the square region $\Omega_{\text{obs}}(z)$ inside the domain Ω such that the influence matrix of h_{obs} matches as closely as possible the influence matrix of the

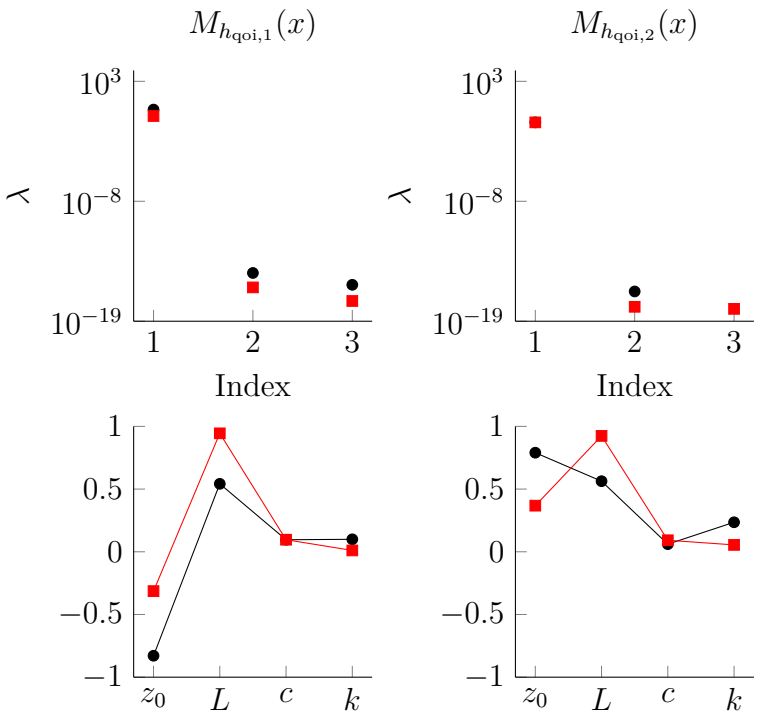


Figure 3.8 Eigenvalues (top row) and eigenvectors (bottom row) associated with the influence matrices for the prediction scenario (black circles) and for the validation scenario (red squares). The left and right columns display the sensitivity indices related to the first and second QoIs, respectively. For the sake of clarity, only the first (and only relevant) eigenvector is displayed for each influence matrix.

QoIs at the validation scenario. Instead of solving directly the optimization problem (3.14), we plot the objective function in (3.14) for various positions of the squared region inside Ω. We actually consider for the squared regions the grid cells shown in Figure 3.6. Figure 3.9 presents the results for both QoIs under their respective optimal validation scenario. The best positions for the sensor are indicated in this figure by the squares in which the value of the objective function is the closest to zero.

For the first QoI, we clearly note that it is best to position the sensor within Ω_1, which is an obvious result since we ultimately seek to predict the mean concentration of pollutant in Ω_1. The optimal region of observation seems to extend to the region where recirculation occurs, as shown in Figure 3.10. To some extent, it also spreads upstream above the first dock and downstream right above the second dock. Moreover, we observe that positioning the sensor in the upper-half of the domain would be useless for the prediction of the first QoI. This result is in fact quite intuitive since, under the optimal validation scenario shown in Figure 3.7, very little pollutant flows into those regions while a certain amount of pollutant indeed reaches Ω_1. Finally, the influence matrices associated with the mean concentration in the regions located

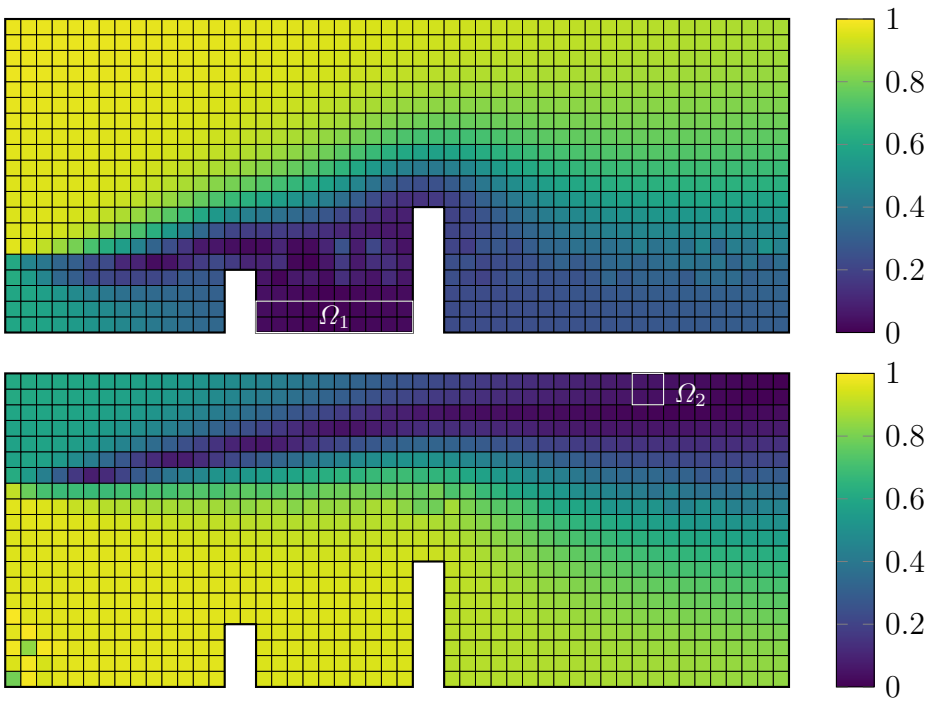


Figure 3.9 Value of the objective function in (3.14) for the first QoI (top) and second QoI (bottom).

after the second dock share some similarities with the influence matrix $M_{h_{qoi,1}}(x_{\text{val}})$, but not as much as the region between the two docks. We also remark that the fluctuations in the objective function in the plume located downstream could be explained by the accumulation of numerical errors resulting from the computation of the gradient (see 3.B).

For the second QoI, it is again optimal to position the sensor inside the region of interest, Ω_2 in this case. The pollutant sensor can also be placed optimally right upstream and right downstream of Ω_2. Interestingly, it does not seem optimal to put the sensor right inside the plume of the pollutant. One should rather probe the boundary region of the plume next to the QoI since variations in the concentration are larger there and the measurements would thus be more sensitive to the location of the pollutant release.

3.5 Conclusion

We have addressed in this paper the issue of designing optimal validation experiments tailored toward the prediction of a QoI. To the best of our knowledge, this problem has been rarely considered in the literature. We thus provide a careful description of a mathematical model, the role of the various input parameters, and the objectives when performing predictions using

the model. We have also carried out a consistent treatment of the uncertainty affecting the model parameters, the experimental and computed observations, and the QoI. Within this framework, we have introduced the influence matrix, computed using the Active Subspace method, as a means to provide a quantitative description of the response surfaces associated with the various model functionals. We have then proposed a methodology to design an optimal validation scenario and identify optimal observations based on the comparison of influence matrices. The methodology essentially consists in the solution of two optimization problems: 1) the first optimal design problem (3.13) allows the computation of the optimal validation scenario; 2) the second optimal problem (3.14) finds the best measurements to be performed on the system of interest.

The methodology was tested on two numerical examples. The results for a simple projectile problem demonstrated that the optimal validation scenario was able to recover the essential features of the influence matrix associated with the QoI at the prediction scenario. The influence matrices obtained with two different observable functionals were compared over a given time period and allowed us to identify an optimal observation functional and sensor parameter. The results for the pollutant transport problem highlighted the fact that the optimal validation experiment does indeed depend on the QoI we wish to predict. They confirmed as well the fact that, in the case of non-observable QoI, one can still identify observation functionals sharing similar influence matrices. This numerical example also shows that the proposed methodology can be viably used for applications of engineering interest.

Several hypotheses and simplifications have been made throughout this work, limiting to some extent the scope of the predictive problem. Alleviating the hypothesis that the control parameters and sensor parameters are not uncertain could lead to more robust validation experiments, especially when these control parameters consist of boundary and/or initial conditions. Also, the use of the Active Subspace method to define the influence matrix necessitates some regularity on the observation and QoI functionals. For models lacking regularity, one could envision the use of the *variogram analysis of response surfaces* [105] as a means to compute the influence matrix.

Moreover, in addition to employing the sensitivity analysis to design validation experiments, it could be used for the analysis of existing data sets. By describing the influence matrix of a model under a data set, one could for instance find specific data within this set that could be utilized for validation purposes. Adversarial validation scenarios aimed at testing a wide range of prediction settings may also be found. In the same vein, it could be insightful to apply the influence matrix to cross-validation approaches.

Finally, methods to quantify the model errors rely on some sort of comparison between the

reality and the model observables. It could be interesting to analyze the impact of the choice of the experiment on the modeling error for the QoIs at the prediction scenario. The proposed methodology could perhaps better inform the modeling error, which we will investigate in a future research.

3.A Velocity Field for the Pollutant Transport Example

The velocity field v for the pollutant transport problem is obtained by solving the Stokes' equation with the following boundary conditions

$$-\nu \Delta v + \nabla \beta = 0, \qquad \text{in } \Omega, \tag{3.23a}$$

$$\nabla \cdot v = 0, \qquad \text{in } \Omega, \tag{3.23b}$$

$$v = 0, \qquad \text{on } \Gamma_{\text{banks}} \cup \Gamma_{\text{dock}}, \tag{3.23c}$$

$$v = v_D, \qquad \text{on } \Gamma_{\text{west}}, \tag{3.23d}$$

$$n \cdot (\nu \nabla v - \beta I) = 0, \qquad \text{on } \Gamma_{\text{east}}, \tag{3.23e}$$

where $\nu \approx 0.001$ is approximately the dynamic viscosity of water and β is the pressure. The Dirichlet conditions v_D consist in the fully developed quadratic flow

$$v_D(z) = v_{\max}(2 - z_2)z_2, \tag{3.24}$$

where the maximal velocity $v_{\max}$ is taken here to be equal to unity. We solve the weak formulation of the problem with the open-source Finite Element code FEniCSx [106]. We discretize the velocity field with Lagrange elements of degree two and the pressure field with Lagrange elements of degree one. The meshes employed will be described briefly in Appendix 3.B. The streamlines of the flow are displayed in Figure 3.10.

We observe the presence of three recirculation zones, one before the first dock, a large one between the two docks, and another small one after the second dock. The choice of the Stokes model to represent the velocity field is only a coarse approximation for this application. However, the sole purpose of this model is to obtain a velocity field for the pollutant transport model. For example, if the velocity field were given by the solution to the Navier-Stokes equations, then the pollutant transport model (3.20) would be different from the one under study. However, we would use the same methodology to obtain the optimal validation scenarios.

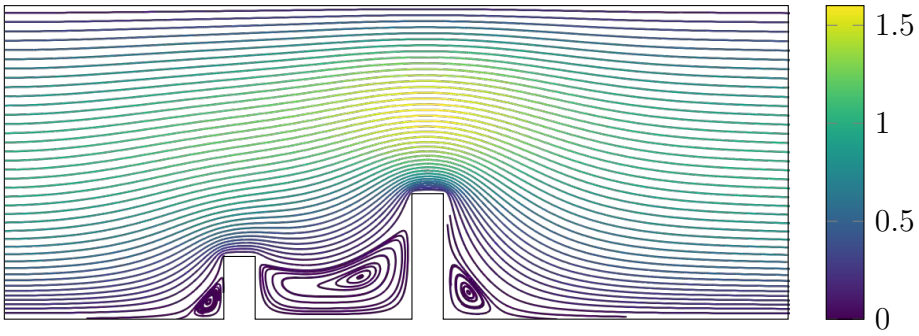


Figure 3.10 Streamline of the velocity field v

3.B Gradient for the Contaminant Transport Example

The computation of the gradient (with respect to the parameters) of the QoI and the observables for the pollutant transport model described in Section 3.4.2 is somewhat involved. In order to do so, we compute the adjoint solution of the diffusion-advection equation (3.20) and use it to obtain the derivatives [107]. As previously mentioned, the state variables (v, β) for the velocity and pressure of the water can be solved independently from the advection-diffusion equation and do not vary with the parameters.

We explicitly introduce a lift function $\bar{\phi}_D$ to account for the Dirichlet boundary condition $\phi = \phi_D$, as we wish to compute the derivative of the QoI and observables with respect to the parameters that define ϕ_D, so $\phi = \bar{\phi} + \bar{\phi}_D$. The weak form of the advection-diffusion reads:

$$\begin{aligned}&\text{Find } \bar{\phi} \in V = \{\psi \in H^1(\Omega);\ \psi|_{\Gamma_{\text{west}}} = 0\} \text{ such that}\\ &\underbrace{\int_\Omega \exp(k)\nabla\bar{\phi}\cdot\nabla\psi + \nabla\cdot(v\bar{\phi})\psi\,\mathrm{d}z}_{=a(\bar{\phi},v)} + \underbrace{\int_\Omega \exp(k)\nabla\bar{\phi}_D\cdot\nabla\psi + \nabla\cdot(v\bar{\phi}_D)\psi\,\mathrm{d}z}_{=a(\bar{\phi}_D,\psi):=l(\psi)} = 0, \quad \forall\ \psi \in V.\end{aligned} \tag{3.25}$$

The equilibrium relation for this problem reads:

$$r(x,\theta,\phi) = a(\phi,\cdot) = a(\bar{\phi},\cdot) + a(\bar{\phi}_D,\cdot) \in V^*, \tag{3.26}$$

where V^* is the dual space of V. From [107, Section 1.6], the computation of the derivative is performed in two steps. We first solve the adjoint problem for λ such that

$$r_\phi(p,\phi(p))^*\lambda = -h_\phi(p,\phi(p)), \tag{3.27}$$

and then compute the gradient according to

$$\nabla_p h(p,\phi) = r_p\Big(p,\phi(p)\Big)^* \lambda + h_p\Big(p,\phi(p)\Big). \tag{3.28}$$

The subscripts ϕ and p stands for the variables with respect to which we perform differentiation and * denotes the adjoint. For the problem at hand, the partial derivative of the equilibrium relation in the direction $\hat{\phi}$ is

$$r_\phi(p,\phi(p))\hat{\phi} = a(\hat{\phi},\cdot), \tag{3.29}$$

while its adjoint is

$$\begin{aligned}\langle r_\phi(p,\phi(p))^*\psi,\phi\rangle_{V^*,V} &= \langle r_\phi(p,\phi(p))\phi,\psi\rangle_{V^*,V}\\ &= a(\phi,\psi)\\ &= \int_\Omega \exp(k)\nabla\phi\cdot\nabla\psi + \nabla\cdot(v\phi)\psi\,\mathrm{d}z.\end{aligned} \tag{3.30}$$

Hence, the adjoint problem reads:

$$\begin{aligned}&\text{Find } \lambda\in V \text{ such that}\\ &\int_\Omega \exp(k)\nabla\psi\cdot\nabla\lambda + \nabla\cdot(v\psi)\lambda\,\mathrm{d}z = -\frac{1}{|\omega|}\int_\omega \psi\,\mathrm{d}z\,, \qquad \forall\,\psi\in V,\end{aligned} \tag{3.31}$$

where ω can be either Ω_1, Ω_2, or Ω_{obs}. We now need to compute the partial derivative of the equilibrium relation $r(p,\phi(p))$ and of the QoI/observables $h(p,\phi(p),z)$ with respect to the parameters p. As a reminder, the control parameters for this problem are the location, width, and height for the Dirichlet condition, that is $x = (z_0, L, c)$), respectively, and the model parameter is the diffusivity term $\theta = k$. It is at this point that the lift function will be employed. By applying the chain rule, we obtain the partial derivative with respect to the control parameters x in the direction $\hat{x}$

$$r_x(p,\phi(p))\hat{x} = r_{\bar{\phi}_D}(p,\phi(p))\bar{\phi}_{D,x}(x,z)\hat{x} = a(\bar{\phi}_{D,x}(x,z)\cdot\hat{x},\cdot), \tag{3.32}$$

where $\bar{\phi}_{D,x}$ denotes the gradient of $\bar{\phi}_D$ with respect to x. Its adjoint is computed as

$$\begin{aligned}\langle r_x(p,\phi(p))^*\psi,\hat{x}\rangle_{\mathcal{X}^*,\mathcal{X}} &= \langle r_x(p,\phi(p))\hat{x},\psi\rangle_{V^*,V} = a(\bar{\phi}_{D,x}(x,z)\cdot\hat{x},\psi)\\ &= \int_\Omega \exp(k)\nabla(\bar{\phi}_{D,x}(x,z)\cdot\hat{x})\cdot\nabla\psi + \nabla\cdot\Big(v(\bar{\phi}_{D,x}(x,z)\cdot\hat{x})\Big)\psi\,\mathrm{d}z\\ &= \int_\Omega \Big(\exp(k)\nabla\bar{\phi}_{D,x}(x,z)\cdot\nabla\psi + \nabla\cdot(v\bar{\phi}_{D,x}(x,z))\psi\Big)\cdot\hat{x}\,\mathrm{d}z.\end{aligned} \tag{3.33}$$

So the gradient of the QoI/observable h with respect to the control parameters is given by

$$\nabla_x h(p,\phi(p)) = \int_\Omega \exp(k)\nabla\bar{\phi}_{D,x}(x,z)\cdot\nabla\lambda + \nabla\cdot(v\bar{\phi}_{D,x}(x,z))\lambda\,\mathrm{d}z + \frac{1}{|\omega|}\int_\omega \bar{\phi}_{D,x}(x,z)\,\mathrm{d}z. \tag{3.34}$$

The derivative with respect to the model parameter $\theta = k$ can be similarly computed as

$$\nabla_\theta h(p,\phi(p)) = \int_\Omega \exp(k)\nabla\phi\cdot\nabla\lambda\,\mathrm{d}z. \tag{3.35}$$

The adjoint solution (3.31), the gradient (3.34), and the derivative (3.35) are evaluated within the Finite Element code FEniCSx [106] using the module UFL [108]. The pollutant concentration ϕ and the adjoint solution λ are discretized with Lagrange elements of degree one.

Several factors have guided the design of the meshes. First, for each value of control parameter x, a new solution for the pollutant concentration ϕ needs to be computed. Second, for each definition of the observable function h_{obs} and QoI functional h_{qoi}, a new adjoint solution λ needs to be computed. Ideally, for each of the control parameters x and QoI/observable functionals considered, one should estimate the discretization errors on the concentration ϕ and on the adjoint solution λ and adapt the mesh accordingly. Since the purpose of this pollutant example is to illustrate the proposed methodology to design validation experiments, we do not adopt this approach to design the meshes. We consider essentially four fixed meshes in this study. All of these meshes are refined near Γ_{west}, near the corners of the docks and inside the recirculation region. The first two meshes are employed to find the validation scenario and they are conformal with respect to Ω_1 and Ω_2, respectively. Once the optimal validation scenario has been computed for each QoI, two new meshes are constructed with additional refinement near the regions where the concentration is high. These two last meshes are employed to find the optimal type of observable. These general and heuristic choices are made

in order to control the discretization errors without employing involved mesh adaptation techniques. An example of such meshes is displayed in Figure 3.11. The meshes possess between 80,000 and 100,000 triangular elements.

3.C Solution of the Optimal Design Problems

The optimization problems (3.13) are solved with the solver NOMAD [74] using the implementation of the NOMAD code [109] in the programming language Julia. The main reason is that NOMAD is a derivative-free solver, meaning that only the value of the objective function is needed. The computation of the derivative of the objective function in problems (3.13) and (3.14) requires computing the Hessian of the observation and QoI functionals h_{obs} and h_{qoi} with respect to the control and model parameters x and θ. Their calculation is often too computationally intensive, so we prefer here to rely on a derivative-free optimization solver. NOMAD allows one to find local minima, but additional resources can be allocated to perform a global search. This option is employed in this work for reasons that are outlined below.

As mentioned in Section 3.3.3, the objective functions in Problems (3.13) and (3.14) may not possess a unique global minimum. Moreover, the objective functions may possess several local minima. Hence, it is beneficial to perform a global search in the controlled environment $\mathcal{X}_{\text{lab}}$ and constrained set $\mathcal{Z}_{\text{lab}}$ to escape the basin of attraction of a local minimum and hopefully find a global minimum.

To gain insight into the behavior of the objective function of Problem (3.13) and to assess the performance of NOMAD, we perform two verification tests. These two verification tests consist in solving (3.13) for the two models described in Sections 3.4.1 and 3.4.2, but for a controlled environment $\mathcal{X}_{\text{lab}}$ now equal to the control space $\mathcal{X}$. A global minimum is guaranteed to exist in that case, which precisely corresponds to the prediction scenario. For the projectile model, the control parameters space is $\mathcal{X} = [0.005, 5] \times [0.005, 0.1] \times [0, 2] \times [10, 120]$. As described earlier, the prediction scenario is $x_{\text{pred}} = (0.05, 0.01, 1, 100)$. We initialize the optimization algorithm with ten different initial points distributed according to the Latin Hypercube algorithm. Table 3.4 reports the local minimum attained for each initial point. The values of the objective function normalized with respect to $\left\| M_{h_{\text{qoi}}}(x_{\text{pred}}) \right\|$ indicate that the quality of all local minima is similar. However, when we look at the error, measured as the l_2 distance between the computed optimal point and the prediction scenario x_{pred}, we observe that no initial point leads to the prediction scenario. This result can be explained by analyzing more closely the optimal points. On one hand, all have succeeded in recovering the value of the mass m, diameter ℓ, and initial velocity v_0 of the prediction scenario x_{pred}.

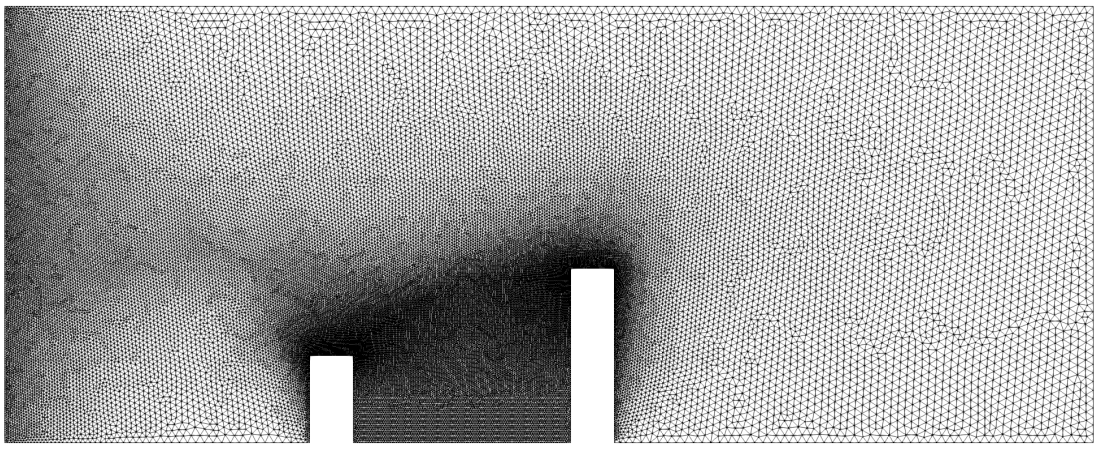


Figure 3.11 Mesh employed to compute the optimal validation scenario for the prediction of the mean concentration inside Ω_1.

On the other hand, values obtained for the initial position u_0 are somewhat different. This result arises from the fact that the QoI (3.15) is affine with respect to the initial position u_0. Hence, the influence matrix is the same whatever the value of the initial position u_0 (for fixed value of the other control parameters), and therefore reflected by the values of the objective function. We may conclude that a unique global minimum does not exist for this problem. However, there exists a one-dimensional linear manifold on which the global minimum can be attained.

We repeat the same verification process for the pollutant transport model. The prediction scenario is $x_{\text{pred}} = (0.75, 0.6, 2)$ and the control parameters space is $\mathcal{X} = [0.1, 1.9] \times [0.005, 1] \times [0.1, 10]$. Table 3.5 provides the solutions obtained using NOMAD for each initial control parameter. The verification experiments number 3, 9, and 10 seem to indicate the existence of a local minima away for the prediction scenario. However, the values of their normalized objective function indicate that the influence matrices of the QoI at the optimal points and at the prediction scenario x_{pred} are similar. The other verification experiments possess lower value of normalized objective function, even if some optimal points are not very close to the prediction point.

These verification experiments demonstrate the robustness of the solver NOMAD for solving the optimal design problem (3.13). We are thus confident that the optimal validation scenarios obtained earlier for the projectile and pollutant transport problems are reasonable local minima of the objective function (3.13) under their respective constrained sets $\mathcal{X}_{\text{lab}}$.

Table 3.4 Results of the ten verification experiments for the projectile problem

	Initial Point				Optimal Point				Error	Normalized Objective Function
	m	ℓ	u_0	v_0	m	ℓ	u_0	v_0		
1	2.54	0.0236	0.880	76.9	0.0500	0.0100	1.13	100	0.126	1.6e-05
2	1.44	0.0245	0.438	57.0	0.0500	0.0100	0.507	100	0.493	1.0e-06
3	0.764	0.0509	0.134	43.2	0.0500	0.0100	1.75	100	0.752	1.0e-06
4	1.72	0.0543	1.29	113	0.0500	0.0100	0.870	100	0.130	1.0e-06
5	4.26	0.0415	0.207	97.8	0.0500	0.0100	0.41	100	0.590	4.0e-06
6	3.57	0.0890	0.699	107	0.0500	0.0100	1.98	100	0.977	1.0e-06
7	4.75	0.0655	1.66	33.4	0.0500	0.0100	1.70	100	0.703	1.0e-06
8	2.15	0.0117	1.06	27.7	0.0500	0.0100	0.0653	100	0.935	1.0e-06
9	0.324	0.0937	1.96	18.7	0.0500	0.0100	1.23	100	0.230	1.0e-06
10	3.14	0.0790	1.49	75.9	0.0500	0.0100	1.61	100	0.606	1.6e-05

Table 3.5 Results of the ten verification experiments for the pollutant transport problem

	Initial Point			Optimal Point			Error	Normalized Objective Function
	z_0	L	c	z_0	L	c		
1	0.975	0.949	5.14	0.745	0.592	1.99	0.0118	1.8e-4
2	1.06	0.802	1.98	0.700	0.523	1.95	0.105	5.2e-4
3	1.28	0.716	6.79	0.979	0.963	2.33	0.539	1.2e-2
4	0.654	0.626	0.950	0.800	0.680	2.06	0.112	2.4e-3
5	0.541	0.510	7.54	0.785	0.655	2.04	0.0772	1.6e-3
6	1.50	0.437	2.97	0.696	0.516	1.95	0.114	1.1e-4
7	1.65	0.344	9.50	0.826	0.720	2.09	0.170	3.9e-3
8	0.323	0.254	8.83	0.756	0.610	2.00	0.0133	2.7e-4
9	1.83	0.111	3.72	0.933	0.894	2.25	0.429	9.8e-3
10	0.128	0.0403	4.26	0.905	0.847	2.21	0.357	8.3e-3

CHAPTER 4 GOAL-ORIENTED CALIBRATION OF A MODEL AND ITS MODELING ERRORS

In Section 2.6, we presented the MLE and MAP estimators for perfect and misspecified model. We have seen that both approaches asymptotically seek the value of the model parameters θ that minimizes the Kullback-Leibler divergence $\mathrm{KL}(Y_{\mathrm{exp}}, Y^{\theta})$. This objective is merely a choice of the calibration functional.

In this chapter, we explore the issue of tailoring the calibration process toward the predictive objective. Our proposed calibration approach is based upon the Kennedy O'Hagan (KOH) framework, but we introduce two important points of departure. First, each value of the model parameters θ generates its own model with corresponding modeling errors. Hence, the modeling errors are now viewed as a function of the model parameters, and a function of the experimental design and data. Second, the calibration of the model parameters is no longer performed via MLE or MAP. We define the calibration objective of θ as the minimization of a functional defined in terms of the modeling discrepancy. This functional can then be tailored toward the predictive objective of the model. We illustrate our goal-oriented calibration approach on three numerical examples. The first example is a simple synthetic example that will be employed to illustrate the various steps of our methodology. Moreover, this example will allow us to emphasize the differences with other calibration approaches. The second example is a manufactured problem, where we assume that the model is perfect. With this example, we verify whether the proposed methodology can recover the true model parameters. The last example studies the thermal problem presented at the Sandia Validation Challenge [1, 42]. Our aim here is not to provide an answer to the challenge, but merely to exploit the available experimental data.

4.1 Introduction

Physical models are merely abstractions of systems of interest. Their construction allows one to better understand or simulate physical phenomena, and possibly perform predictions to guide a decision-making process. The modeling act necessarily involves simplifying hypotheses to render the physical system of interest intelligible. Such hypotheses introduce two important ideas that are intricately related:

1. the simplifying hypotheses inevitably induce some modeling errors, a notion which is expressed in the first part of the famous quote by George Box "All models are wrong,

but some are useful" [82, p.61]. However, the second part of the statement conveys the idea that models can still be employed to perform predictions as long as the modeling errors are controlled for these predictions;

2. the modeling hypotheses usually introduce unknown parameters, usually arising from phenomenological closure models. The model parameters can be regression coefficients, can represent material properties, or be introduced as parameters of a statistical distribution, etc. Since such model parameters are often not well-known, they need to be calibrated against experimental data gathered by making observations of the system of interest. However, in some cases, these parameters may only make sense in controlled scenarios far away from the prediction one.

Values of the model parameters directly affect the modeling errors in the sense that different values may yield different model predictions. Since no model is perfect, the idea that there exists a true value for the model parameters is not always well founded. By calibrating the model parameters, one does not seek its true value, but rather an optimal one. The definition of optimality is described, either implicitly or explicitly, in the choice of the formulation of the calibration problem. This assessment, in conjunction with the inevitable presence of modeling errors, forms the basis of this work.

Our vision of the calibration problem is inspired by numerous works. First and foremost, we mention the seminal paper of Kennedy and O'Hagan [18]. Here, we will refer to the framework proposed in [18] as the KOH framework or the KOH approach, which introduces a way to explicitly incorporate the modeling discrepancy into the calibration process by representing them in terms of Gaussian processes. In this work, we will distinguish the modeling errors and the modeling discrepancy. The modeling errors aggregate the phenomena not captured by a model. The modeling discrepancy is the manifestation of these modeling errors on the observables, i.e., the difference between the model prediction and the experimental data not explained by observation noise. Kennedy and O'Hagan mention in [18, Section 4.3] that the notion of true model parameters should be understood as the model parameters that best fit the data, given a model, a representation of the modeling discrepancy, and observation errors. We will revisit the KOH framework in further detail in Section 4.2.4, since we will adopt their approach to represent modeling discrepancy. Arendt et al. [96] discuss some identifiability issues associated to the KOH framework. Indeed, simultaneously calibrating the model parameters and the modeling discrepancy can mix up their respective influences. The necessity to differentiate the influence of the model parameters and that of the modeling discrepancy constitutes a central component of our work.

The work of Leoni et al. [21] is based upon the KOH framework, but the authors adopt a

different point of view regarding the calibration of the model parameters and the modeling discrepancy. Instead of calibrating simultaneously the model parameters and the modeling discrepancy, they acknowledge that each value of the model parameters defines its own model. Hence, each value of the model parameters generates distinct modeling discrepancy that should be calibrated against experimental data. More precisely, a maximum posterior estimator of the parameters of the modeling discrepancy is computed for each value of model parameters. Our proposed methodology adopts this same point of view regarding the interactions between the model parameters and the modeling discrepancy.

The article by Jiang et al. [20] presents a useful review of calibration methods based on the KOH approach as well as calibration methods framed as optimization methods. They mention that, with optimization-based calibration approaches, one could a posteriori recover the modeling discrepancy in the KOH fashion. We take inspiration of their discussion in [20, Section 2.4] in the development of our goal-oriented methodology. The authors also present a novel calibration approach that recognizes that the definition of the likelihood depends directly upon our assumption regarding the modeling errors. They in fact propose a sequential method to calibrate the model parameters by weighting the contribution of the prior and the updated posterior.

The paper of Hu et al. [22] frames the calibration process as an optimization problem where the discrepancies between the empirical cumulative distribution and the model distribution are minimized. Their notion of optimality is based upon the area validation metric [30]. Their work takes into account aleatory uncertainty affecting the parameters describing the calibration scenarios. Joseph and Melkote [23] proposes to calibrate the model parameters using a least squares formulation. If their calibrated model is not deemed adequate, they correct it with an affine transformation of the model output. If their new model is still not valid, they further add (non-linear) regression terms to decrease the discrepancy with the experimental data.

Oden and Prudhomme [39] propose to perform two Bayesian calibrations of a model. To this end, a set of experimental data is separated into two distinct sets, a calibration set and a validation set. The first calibration employs only the calibration data set, while the second calibration is performed with the calibration and validation sets. For both calibrated models, the distribution of the quantity of interest (QoI) at the prediction scenario is computed and compared with some metrics. If the discrepancy is high, this indicates that the prediction of the QoI is sensitive to the calibration process. In turn, this may hint at the presence of significant modeling errors. Morrison et al. [40] apply this methodology in the context of cross-validation. Instead of considering only one partition of the experimental data, they

consider all possible partitions, which allows the identification of the partition that generates the most discrepancy between the observables and the predictions from the model.

In light of this discussion, and to the best of our knowledge, there are two key aspects of the calibration problem that are scarcely covered in the literature:

1. the predictive objective when using a given model is rarely built in explicitly within the formulation of the calibration problem;

2. for optimization-based formulations of the calibration problem, the modeling errors are not a priori taken into account.

Hence, the objective of this paper is to develop a new calibration process for which one can explicitly control the definition of the objective function while consistently integrating the modeling errors. In other words, having a greater control by defining the objective function in terms of the quantity of interest could potentially allow the modeler or the user to tailor the calibration process with respect to the predictive objective of the model. As discussed previously, the inevitable presence of modeling errors may diminish the quality of the prediction. Thus, it seems logical to formulate the calibration problem so as to reduce some features of the modeling errors, e.g. features relevant to the predictions. For instance, if one wants to employ a model at a specific prediction scenario, the calibration process should choose the value of the model parameters that minimizes the modeling errors at this specific scenario. On the other hand, if the model is to be used over a set of scenarios, then the choice of the model parameters should minimize the modeling errors over this set. In this work, we will consider the modeling errors affecting the observables, not necessarily those affecting the QoIs. Nevertheless, possible avenues to adapt the proposed methodology to QoIs will be made as a conclusion remark.

The paper is organized as follows. Section 4.2 introduces the various notations and concepts necessary to the formulation of the model calibration problem. The precise definition of a model, its modeling errors, and their respective parameters, are laid out. Details about the experimental data are also presented in this section. Section 4.3 describes some well-known calibration processes, namely the Maximum Likelihood Estimation (MLE) and the Bayesian Inference. Both methods are presented with and without the presence of modeling errors. Section 4.4 describes the proposed goal-oriented approach to the calibration problem. Two procedures to calibrate the modeling discrepancy are enunciated. Then, the goal-oriented calibration method is presented. Finally, numerical examples are studied in Section 4.5. First, a simple synthetic example is considered to compare the two methodologies to calibrate the modeling discrepancy. This example is also used to study several goal-oriented calibration

objectives. The second example is a manufactured problem that excludes any modeling errors. This example serves as a sanity check, where we verify that the goal-oriented approach calibrates the model parameters to the value that was used to generate the experimental data. Finally, we consider the thermal problem formulated at the Validation Challenge Workshop [1, 42] held at Sandia National Laboratories in 2006. This example is employed to further illustrate the versatility of the goal-oriented calibration on a more realistic application.

4.2 Preliminary Notions about Predictive Modeling and Modeling Errors

In this section, we introduce various concepts and notations surrounding a model and its associated modeling errors. We first provide a definition of a deterministic model. A particular emphasis will be put on the role each type of parameters plays in this definition. Then, we will recall the various approaches adopted to assess the modeling errors.

In this paper, we shall consider the probabilistic point of view for the treatment of uncertainties and errors. We refer the interested reader to the book of Klir [43] for other characterizations of the uncertainty. We adopt the convention that a random variable will be denoted by an upper-case and that a realization of this random variable will be denoted by a lower-case. Moreover, we will refer to a stochastic process using an upper-case as it should be clear from the context whether it is a random variable or a stochastic process. For instance, we will note $E(\chi)$ the random variable arising from the stochastic process E at the point χ. For the sake of clarity in the presentation, we will omit the underlying probability space (and sample space Ω), apart from specifying a realization of the stochastic process by $E(\cdot, \omega)$, for $\omega \in \Omega$.

4.2.1 Illustrative Model

In order to illustrate the various concepts introduced in this section, we will consider the thermal problem presented at the Model Validation Challenge Workshop by Sandia Laboratories [1,42]. This problem consists in modeling the temperature u inside a solid subjected to a heat flux in one dimension and in assessing whether the temperature in the solid exceeds a certain threshold at a specific time. The proposed model is expressed in the following form

$$u(q, L, u_0, k, \rho C_p, x, t) = u_0 + \frac{qL}{k}\left[\frac{kt}{\rho C_p L^2} + \frac{1}{3} - \frac{x}{L} + \frac{x^2}{2L^2} - \frac{2}{\pi^2}\sum_{n=1}^{25}\frac{1}{n^2}\exp\left(-n^2\pi^2\frac{kt}{\rho C_p L^2}\right)\cos\left(n\pi\frac{x}{L}\right)\right], \tag{4.1}$$

where u is the temperature function that depends on the prescribed heat flux q, the length of the solid L, the initial temperature u_0, the thermal conductivity k, the volumetric capacity ρC_p, the position x in the solid, and the time t. Beside the model, the authors also provide experimental data recorded at various conditions with several replicates. It is mentioned in the paper that the data are in fact synthetic data, but we will treat them as if they were real experimental data. We will make reference to this thermal model and these experimental data throughout this section. Moreover, the problem will be revisited in Section 4.5.3 for the purpose of a numerical experiment.

4.2.2 Abstract Model of the System of Interest

We define a deterministic model as a set of equations r of the form

$$r(p, \theta, u) = 0, \tag{4.2}$$

that describe the relations between the *control parameters* $p \in \mathcal{P}$ of finite dimension, the *model parameters* $\theta \in \mathcal{T} \subseteq \mathbb{R}^{d_\theta}$, and the *state variable* $u \in \mathcal{U}$, where $\mathcal{U}$ is assumed to be a Banach or a Hilbert space. These relations r encode all hypotheses and assumptions about the system of interest. They may represent first principles, closure models, initial and boundary conditions, and so on. We make the assumption that this model is well-posed, meaning that for each tuple $(p, \theta) \in \mathcal{P} \times \mathcal{T}$, there exists a unique and stable $u \in \mathcal{U}$ satisfying (4.2). This assumption implies the existence of a function $g : \mathcal{P} \times \mathcal{T} \to \mathcal{U}$ such that $g(p, \theta) = u$.

The control parameters $p \in \mathcal{P}$ consist in parameters one can control to influence the system of interest. They can be viewed as the inputs that define the conditions or scenarios upon which the system of interest is subjected. Examples of such control parameters are initial and boundary conditions, geometric parametrizations, and directly observable physical quantities influencing the conditions of the system of interest. The remaining parameters necessary for the well-posedness of the model are the model parameters $\theta \in \mathcal{T}$. They are coined this way because they represent parameters stemming from the various modeling hypotheses. A good example of such a model parameter is the thermal conductivity k: it is not a quantity directly observable, but rather a quantity arising from heat transfer theory. To summarize the difference between these two types of parameters, the control parameters p describe the *scenario* upon which the system of interest is subjected to and the model parameters θ are those that need be inferred through prior information or a calibration process. A last type of parameters consists of the *sensor parameters* $z \in \mathcal{Z}$ that represent the domain of definition of the state variable u. They can be defined as the positions in time and space where one will put a sensor or probe to inspect the system of interest.

For the thermal model (4.1), the control parameters are given by $p = (q, L, u_0)$ since they describe the experiment to be performed. The model parameters are $\theta = (k, \rho C_p)$, the state variable is the temperature u and the equilibrium relation can be obtained by rearranging equation (4.1). The sensor parameters are the pair $z = (x, t) \in [0, L] \times \mathbb{R}^+$ representing the position x where measurements are taken at time t. To obtain this model, several hypotheses have been put forth, such as:

- the solid is assimilated as a one-dimensional entity;
- the solid is subjected to a constant initial temperature field u_0;
- the thermal conductivity k and the volumetric capacity ρC_p are assumed constant throughout the solid.

Once the set of equations (4.2) is solved, the solution of the model can be employed to compute several quantities. First and foremost, one can compute some observables y. As is implied by their names, these quantities are the equivalent of experimental data gathered on the system of interest. These observables are abstracted in terms of the function

$$h\Big(p, \theta, u(z)\Big) = h\Big(p, \theta, g(p, \theta)(z)\Big) = h(p, \theta, z), \tag{4.3}$$

where a slight abuse of notation is made on the last equality in order to simplify the presentation. Another important output of the model is the quantity of interest $h_{\text{qoi}} : \mathcal{P} \times \mathcal{T} \to \mathbb{R}$. It is the QoI that guides the stakeholder of the model in their decision-making process.

For the sake of clarity and conciseness, we will aggregate the control parameters p and the sensor parameters z into the parameters $\chi = (p, z) \in \mathcal{P} \times \mathcal{Z} = \mathcal{X} \subseteq \mathbb{R}^{d_\chi}$. This choice is justified by the fact that we will mainly be interested in the observations of the system of interest. We will thus write the observable function as

$$h(\chi, \theta) = h(p, \theta, z). \tag{4.4}$$

4.2.3 Experiments on the System of Interest

An experimental design will be described by $\chi_{\text{exp}} = (\chi_1, \ldots, \chi_m) \in \mathbb{R}^{d_\chi \times m}$ while the data experiment will be described by $y_{\text{exp}} = (y_1, \ldots, y_m) \in \mathbb{R}^m$. The integer m refers to the number of experimental data gathered during the experiment. Without loss of generality, we will consider that there is only one type of observables recorded at each experiment, i.e., there is only one observation functional h. If a collection of experiments is considered, we will

write the i-th experimental design and experimental data by $\chi_{\text{exp},i}$ and $y_{\text{exp},i}$, respectively. This index will also be propagated to each element of the experimental design and each experimental data, e.g. $\chi_{\text{exp},i} = (\chi_{i,1}, \ldots, \chi_{i,m_i})$.

Table 4.1 Subset of the data from the thermal problem. Taken from Table 1 in [1]

	$\chi = (1000, 0.0127, 25, 0, t)$				$\chi = (1000, 0.0254, 25, 0, t)$			
t	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Exp 6	Exp 7	Exp 8
0.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0
100.0	105.5	109.0	96.3	111.2	99.5	106.6	96.2	101.3
200.0	139.3	143.9	126.0	146.9	130.7	140.4	126.1	133.1
300.0	165.5	170.5	148.7	174.1	154.4	165.9	148.7	157.2
400.0	188.7	193.5	168.5	197.5	174.3	187.2	167.7	177.2
500.0	210.6	214.6	186.9	219.0	191.7	205.8	184.3	194.8
600.0	231.9	234.8	204.6	239.7	207.3	222.4	199.3	210.6
700.0	253.0	254.6	222.0	259.9	221.7	237.6	213.0	225.0
800.0	273.9	274.2	239.2	279.9	235.0	251.7	225.7	238.4
900.0	294.9	293.6	256.4	299.9	247.6	264.9	237.6	251.0
1000.0	315.8	313.1	273.5	319.8	259.5	277.4	248.9	262.9

Let us consider a subset of the experimental data extracted from the thermal challenge [1] and recorded in Table 4.1. From this subset of data, we observe that there are $N = 8$ experiments, each of which contains $m_1 = \cdots = m_8 = 11$ data. These experiments are described for $j = 1, \ldots, 11$ by $\chi_{i,j} = (1000, 0.0127, 25, 0, 100(j-1))$ for $i = 1, \ldots, 4$ and $\chi_{i,j} = (1000, 0.0254, 25, 0, 100(j-1))$ for $i = 5, \ldots, 8$.

The underlying stochastic process Y_{exp} that generated the data y_{exp} is not accessible and never will be. It is the whole objective of the model to approximate this process. Since our model consists in an approximation of the system of interest, we need some additional terms to correct the model. In this work, we adopt the KOH approach [18] that relates the experimental data y_i and the model observables h as

$$y_i = h(\chi_i, \theta) + e_i + \phi_i, \tag{4.5}$$

where e_i is the modeling discrepancy and ϕ_i is the observation error associated to the experimental data y_i obtained at χ_i. Within the probabilistic framework, we can interpret the various terms in (4.5) as realizations of their respective counterpart

$$Y_i^\theta = h(\chi_i, \theta) + E_i + \Phi_i. \tag{4.6}$$

Throughout this work, we will note $Y^\theta(\chi)$ (Y^θ) the random variable (stochastic process) obtained for the value of the model parameters θ specified as superscript. As it will become clearer later, we employ this notation to emphasize the dependence upon θ of the description of such random quantities. The KOH framework provides a way to describe the random variables Y_i, E_i, and Φ_i via the use of stochastic processes. More precisely, the random variables Y_i, E_i, and Φ_i are distributed according to the random variables $Y(\chi_i)$, $E(\chi_i)$, and $\Phi(\chi_i)$, respectively. These last random variables originate from the Gaussian processes Y, E, and Φ, and are related via the KOH relation

$$Y^\theta(\chi) = h(\chi, \theta) + E(\chi) + \Phi(\chi). \tag{4.7}$$

It is important to note that the assumptions about the dependence structure between the random variables Y_i and Y_j should be translated into the dependence structure of $Y(\chi_i)$ and $Y(\chi_j)$, hence specifying the covariance kernel of Y. By considering the additional term of the modeling discrepancy E, we hope that the Gaussian process Y can now reasonably match the underlying stochastic process Y_{exp}.

We now describe in greater detail the stochastic processes Y_{exp} and Φ while postponing the discussion about the modeling discrepancy E to the next subsection. The true underlying stochastic process Y_{exp} depends on the parameters χ. A major problem with Y_{exp} is the fact that we only have access to limited information about it. This limited information consists in the realizations of Y_{exp} encoded in the data y_{exp}. One could envision the collection of more data, but the precise description of the stochastic process Y_{exp} will still be elusive for two reasons:

1. We cannot repeat the experiment described by χ_{exp} an infinite number of times. Hence, we can only approximate $(Y_{\text{exp}}(\chi_1), \ldots, Y_{\text{exp}}(\chi_m))$ by an empirical distribution.

2. The parameter space $\mathcal{X}$ is usually not finite. It is then impossible to observe a realization of the stochastic process for all $\chi \in \mathcal{X}$. Again, we can only approximate $Y_{\text{exp}}(\cdot, \omega)$ by performing some experiments at the experimental design χ_{exp}.

The fact that the available information on Y_{exp} is scarce will influence our discussion on the modeling discrepancy.

For the thermal problem example, we can assume that the various experiments are independent. Moreover, we can assume that the random variables $Y_{i,j}$ and $Y_{i,j'}$ are correlated since they are taken from the same experiment at different times. The correlation should thus be defined with respect to the time parameter t. Since the other parameters do no vary within

an experiment and the experiments are considered independent, it is not possible to infer a dependence structure with respect to the other parameters.

The observation error Φ is directly related to the experimental apparatus/protocol employed to record the experimental data. It corresponds to the measurement uncertainty, usually provided by the sensor manufacturer. We will assume that the observation noise is described by a Gaussian process $\Phi \sim \mathcal{GP}(m_\Phi(\chi), k_\Phi(\chi, \chi'))$, with mean function m_Φ and kernel function k_Φ. It is typically assumed that the observation error consists of a white noise, meaning that $m_\Phi(\chi) = 0$ and that $k(\chi, \chi') = \sigma^2(\chi)\delta(\chi, \chi')$, where δ denotes the delta Dirac function. This particular choice for the Gaussian process stems from the assumption that the random variables $\Phi(\chi_i)$ and $\Phi(\chi_j)$ are independent whenever $i \neq j$. The random variable $\Phi(\chi)$ possesses a normal distribution with zero mean and prescribed variance $\sigma^2(\chi)$. The variance σ^2 is often considered constant (the homoscedastic case), but it may not always be (the heteroscedastic case).

In the case of the thermal problem, the authors mentioned that there is no observational noise affecting the experimental data. However, for the sake of our presentation, we will assume that there is a small observation errors by choosing $\Phi \sim \mathcal{GP}(0, k(\chi, \chi'))$, with $k(\chi, \chi') = 0.01^2\delta(\chi, \chi')$ taken to be small with respect to the quantities reported in Table 4.1.

4.2.4 Modeling Errors

We now describe more precisely the nature and significance of the modeling discrepancy E. From Equation (4.7), the modeling discrepancy E can be viewed as the additive correction that we apply to the model observation h and observation error Φ in order to recover the experimental stochastic process Y_{exp}. Again, the modeling discrepancy is the manifestation of the modeling errors in the observables. The modeling errors aggregate the errors generated by the various hypotheses and assumptions employed to define the model and, to a lesser extent, the observation errors. Thus, the modeling errors can stem from the following sources:

- neglecting phenomena that significantly affect the system of interest, phenomena that may have been missed in the formulation of the equilibrium relation (4.2) or the observation functional h;
- ignoring the presence of *hidden* control parameters, which could greatly influence the system of interest, but are not accounted for in the model;
- mischaracterizing the regime or the dominant phenomena in the system of interest. An example is assuming that a phenomenon is linear whereas it is non-linear;

- misspecifying the observation error Φ. The modeling discrepancy E must then capture these misspecifications;
- the choice of the model parameters θ. We will come back to this particular point later in this section.

As already mentioned, one has only fragmentary information regarding the experimental stochastic process Y_{exp}. In turn, it implies that one has limited information regarding the modeling discrepancy stochastic process E. Even if the information about the modeling discrepancy is limited, it is desirable to provide a representation, and this for several reasons. For example, one may want to have an approximation of the modeling discrepancy over $\mathcal{X}$ for interpolation purposes. One can then identify some regions of the control parameter space $\mathcal{X}$ in which the model behaves adequately or poorly. Trends of the modeling discrepancy can also be extracted to guide the process of model improvement. One can also want to extrapolate the modeling discrepancy to regions of $\mathcal{X}$ where no experimental data have been collected. This case can occur when one wants to make predictions at a future time or for conditions that cannot be produced in a controlled environment. It is important to note that the assumptions made about the correlations and independence of the experimental data must also be translated to the modeling discrepancy. This is merely a choice, these assumptions could also be incorporated into the observation errors, but here we decide to restrict the meaning and purpose of the observations errors Φ.

We represent the modeling discrepancy as a Gaussian process $E^\xi \sim \mathcal{GP}(m(\chi;\xi_m), k(\chi,\chi';\xi_k))$ with mean function m and kernel function k, as done in the original work of Kennedy and O'Hagan [18]. The Gaussian process E^ξ that describes the modeling discrepancy is parametrized by the parameter $\xi \in \mathcal{E}$. The parameter $\xi = (\xi_m, \xi_k)$ consists of parameters $\xi_m \in \mathcal{E}_m$, specifying the mean function m, and parameters $\xi_k \in \mathcal{E}_k$, specifying the kernel function k. We adopt this representation since it allows the modeling of a vast family of functions via its kernel [69]. It is a property we indeed seek since we would like to capture the part of the experimental data that cannot be represented by our model. The KOH relations now read

$$Y^{\theta,\xi}(\chi) = h(\chi,\theta) + E^\xi(\chi) + \Phi(\chi). \tag{4.8}$$

Another important assumption made in this work is that the modeling discrepancy E is independent of the observation errors Φ. We introduce the *residuals*, defined as

$$r_i(\theta) = y_i - h(\chi_i,\theta), \tag{4.9}$$

which we will often allude to in this paper. The residuals depend upon the model parameters θ since they amount to the difference between the experimental data and the model observables.

Again, let us consider the subset of experimental data of the thermal problem of Section 4.2.1 displayed in Table 4.1. We compare these experimental data with the model observables for model parameter $\theta = (k, \rho C_p)$, with $k = 0.055$ and $\rho C_p = 4\times10^5$. From the residuals, we want to construct a representation of the modeling discrepancy E based on the relation (4.8). We consider, for the sake of illustration, the following representation of the modeling discrepancy

$$E^{\xi} \sim \mathcal{GP}(m(\chi;\xi_m), k(\chi,\chi');\xi_k),$$

where

$$\begin{aligned} m(\chi;\xi_m) &= \xi_{m,1} + \xi_{m,2}L + \xi_{m,3}t, \\ k(\chi,\chi';\xi_k) &= \xi_{k,2}\exp\left(-\frac{1}{2}\frac{(t-t')^2}{\xi_{k,1}^2}\right). \end{aligned}$$

This specific choice of covariance kernel reflects our assumptions about the dependence of the experimental data. First, we have experimental data at two different values of parameter L and for multiple instances of time t, so we can capture a linear mean of the modeling discrepancy. Within an independent experiment, only data for varying time t are available, so the kernel depends only upon this parameter. We employ a radial basis kernel as the modeling discrepancy is supposed to be smooth here. Figure 4.1 shows both the resulting stochastic process Y and the Gaussian process E^{ξ} for $\xi_m = (13.6, -629, 0.00149)$ and $\xi_k = (188, 131)$.

From Figure 4.1, we observe the scarcity of information about the system of interest. First, we have available some data only at times $t \in \{0, 100, \ldots, 900, 1000\}$, meaning that the behavior of the system at other instances can only be inferred with the use of the model. Moreover, even at times $t \in \{0, 100, \ldots, 900, 1000\}$, we do not know the precise behavior of the system of interest since only four replicates are available. We can also note the difference between the replicates, denoting an intrinsic variability in the system of interest. It implies that the deterministic model cannot capture this variability, and since the observation error Φ is considered negligible, it is the modeling discrepancy E that should account for this randomness in the observations.

At this point, it is important to emphasize that the *modeling* process (4.2)–(4.4) and the modeling discrepancy E are very different in nature. On one hand, the model (4.2) originates from abstracting the system of interest with given hypotheses and assumptions. On the other hand, the modeling discrepancy consists in an empirical model, based on some assumptions

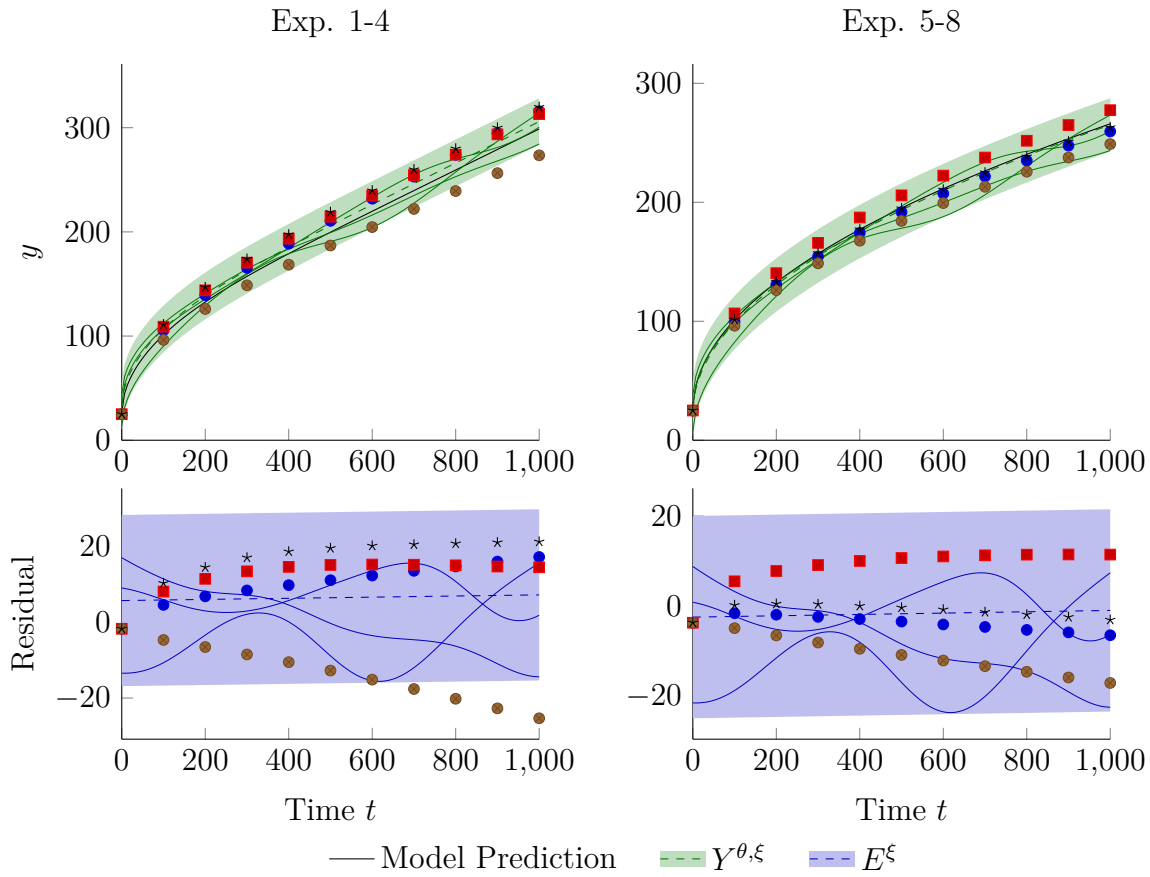


Figure 4.1 Illustration of the Gaussian process $Y^{\theta,\xi}$ (top row) and the modeling discrepancy E^{ξ} for the thermal problem. The experimental data and the corresponding residuals are displayed by various marks. The mean functions of the Gaussian processes are displayed in dashed and some realizations are shown with thin lines.

about the experimental data.

4.2.5 Another Representation of the Modeling Errors

The previous discussion described but one way to account for the modeling errors via the approximation of the modeling discrepancy E. Indeed, several contributions have been devoted to the study of the modeling errors and different approaches have emerged to characterize them. Among those, we mention the work of Sargsyan et al. [25]. The main idea proposed in these works is to make the model parameters θ random (now denoted by Θ) and to compute their distribution such that

$$Y(\chi) \approx h(\chi, \Theta) + \Phi(\chi). \tag{4.10}$$

It is important to note that there is no modeling discrepancy E introduced in the above equation. It is in fact the variability of the model parameters Θ that explains, through h, the experimental stochastic process Y_{exp}. The calibration process now consists in determining the parameters in the distribution of Θ. This approach has the advantage that the propagation of the modeling errors to the QoI q is now straightforwardly computed as $Q = h_{\text{qoi}}(\chi, \Theta)$. However, this approach modifies the nature of the uncertainty affecting Θ. The assumption that the model is deterministic implies that the model parameters θ take one value. We can be uncertain about its value, but this uncertainty is purely epistemic in nature, since, given our definition of the calibration, an increasing set of experimental data will reduce the uncertainties in Θ (pending that they are identifiable). However, transferring the modeling errors, which is essentially a mixture of aleatory and epistemic uncertainty, into the model parameters Θ has the effect of changing its definition and the interpretation of its associated uncertainty. This approach can be viewed as a modification of the modeling assumption, where we transform the deterministic model parameter θ into a random variable Θ.

Another approach taken by Oliver et al. [26] acknowledges the fact that the modeling errors originate from the inability of the model to capture some features of the system of interest. They proposed to enhance the original model (4.2) by considering the following set of equations $\tilde{r}$

$$\tilde{r}(\chi, \theta, A, \tilde{u}) = 0. \tag{4.11}$$

The stochastic term A pertains to the additional physics considered in the enhanced model. The new state variable is denoted by $\tilde{u}$. In some sense, the new set of equations $\tilde{r}$ should capture all modeling errors arising from the original model. This new set of equations $\tilde{r}$ can consist in additive or multiplicative corrections to the original set of equations r. The additional physics can also be coupled to the original model. The new calibration problem aims at identifying the model parameters θ and A. The modeling errors can also be straightforwardly propagated to the QoIs. One drawback of this approach is that the process of modifying the equilibrium relation is intrusive, i.e., one needs to modify the solver employed to solve the original equilibrium relation in order to be able to solve the enhanced equilibrium relation (4.11).

In light of this discussion, we decide to adopt the KOH approach encapsulated by the relation (4.8) for the following reasons. First, it provides a non-intrusive way of obtaining a representation of the modeling errors compared to the approach of Oliver et al. [26]. Second, it allows a representation of the modeling errors that cannot be captured by the structure of the model. This is in contrast with the approach that renders the model parameter θ aleatory,

since the resulting stochastic process only consists of an ensemble of model predictions.

In the next section, we address the problem of calibrating the model parameters θ. Since they influence directly the prediction of the model, they in turn affect the modeling discrepancy E. In this sense, a poorly chosen value of the model parameters θ may increase the resulting modeling errors. On the other hand, if values of the model parameters are reasonably evaluated with respect to the experimental data, then the modeling errors would decrease.

4.3 Calibration of Model Parameters and Modeling Errors

We begin this section by briefly describing two ubiquitous calibration approaches, namely the Maximum Likelihood Estimation (MLE) and the Bayesian approach, in the absence of modeling errors. Then, we revisit these two methods within the KOH framework (4.8). Our objectives are twofold. First, we want to highlight the underlying objective of each calibration method. Second, we want to contrast these calibration objectives with that of computing some predictions using the same model.

4.3.1 Calibration in the Absence of Modeling Errors

We assume here that the model h is a perfect representation of the system of interest, that is, devoid of modeling errors. We consider the stochastic process Y^θ

$$Y^\theta(\chi) = h(\chi, \theta) + \Phi(\chi). \tag{4.12}$$

Since the model is assumed perfect, this implies the existence of *true* model parameters θ_0 such that $Y_{\text{exp}} = Y^{\theta_0}$. The objective of the calibration process can then be recast as that of finding the true value θ_0 of the model parameters.

Central to both the MLE and Bayes methods is the likelihood function

$$L(\theta) = \prod_{i=1}^{N} f_{Y \,|\, \theta, \chi_{\text{exp},i}}(y_{\text{exp},i}), \tag{4.13}$$

where $f_{Y \,|\, \theta, \chi_{\text{exp},i}}$ describes the conditional density function of observing the experimental data $y_{\text{exp},i}$ given the model parameters θ and the experimental design $\chi_{\text{exp},i}$. We assume here that the N experiments are independent, hence explaining the product structure of the likelihood. The relation (4.12) and the correlation structure of the observation errors both determine the likelihood function. The Maximal Likelihood Estimation (MLE) consists in

computing the model parameters θ such that

$$\theta^* \in \underset{\theta \in \mathcal{T}}{\operatorname{argmax}}\, L(\theta). \tag{4.14}$$

In other words, one searches for the model parameters θ that maximize the conditional probability of the observables, given the model parameters θ and the experimental designs $\chi_{\text{exp},i}$. Under various hypotheses regarding the model and the observation errors, the MLE problem (4.14) can be cast in other forms. Here, we assume that the observation errors can be described with a Gaussian process. We will note the covariance matrix $\Sigma_{\Phi,i}$ as

$$\Sigma_{\Phi,i} = \begin{bmatrix} k_\Phi(\chi_{i,1},\chi_{i,1}) & k_\Phi(\chi_{i,1},\chi_{i,2}) & \cdots & k_\Phi(\chi_{i,1},\chi_{i,m_i}) \\ k_\Phi(\chi_{i,2},\chi_{i,1}) & k_\Phi(\chi_{i,2},\chi_{i,2}) & \cdots & k_\Phi(\chi_{i,2},\chi_{i,m_i}) \\ \vdots & \vdots & \ddots & \vdots \\ k_\Phi(\chi_{i,m_i},\chi_{i,1}) & k_\Phi(\chi_{i,m_i},\chi_{i,2}) & \cdots & k_\Phi(\chi_{i,m_i},\chi_{i,m_i}), \end{bmatrix} \tag{4.15}$$

With this notation, the likelihood reads

$$L(\theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^{m_i}|\Sigma_{\Phi,i}|}} \exp\left(-\frac{1}{2}\|y_{\text{exp},i} - h(\chi_{\text{exp},i},\theta)\|^2_{\Sigma^{-1}_{\Phi,i}}\right), \tag{4.16}$$

where $h(\chi_{\text{exp},i},\theta)$ represents the tuple $(h(\chi_{i,1},\theta),\ldots,h(\chi_{i,m_i},\theta))$. To simplify the notation, we write $\|v\|^2_W = (W^{1/2}v, W^{1/2}v) = v^T W v$ for $v \in \mathbb{R}^m$, where W is a given positive definite matrix of size m. Under this ubiquitous hypothesis, the MLE problem is equivalent to the weighted least squares problem

$$\theta^* \in \underset{\theta \in \mathcal{T}}{\operatorname{argmin}}\, \frac{1}{2}\sum_{i=1}^{N} \|y_{\text{exp},i} - h(\chi_{\text{exp},i},\theta)\|^2_{\Sigma^{-1}_{\Phi,i}}, \tag{4.17}$$

where the weights associated with the residuals are given by the covariance matrices $\Sigma_{\Phi,i}$.

An alternative approach is the so-called Bayesian approach. This method assimilates the model parameters θ as a random variable Θ for which we have a prior knowledge encapsulated in the density f_Θ. The prior density f_Θ is then updated with the likelihood to obtain the posterior density $f_{\Theta|Y_{\text{exp}}=y_{\text{exp}},\chi_{\text{exp}}}$

$$f_{\Theta|Y_{\text{exp}}=y_{\text{exp}},\chi_{\text{exp}}}(\theta) = \frac{L(\theta)\, f_\Theta(\theta)}{\int L(\theta')\, f_\Theta(\theta')\, \mathrm{d}\theta'}. \tag{4.18}$$

Often, one is interested in the values of the model parameters that are the most probable,

which leads to solving the problem

$$\theta^* \in \underset{\theta\in\mathcal{T}}{\operatorname{argmax}}\, f_{\Theta|Y_{\exp}=y_{\exp},\chi_{\exp}}(\theta), \tag{4.19a}$$

$$\Leftrightarrow \qquad \theta^* \in \underset{\theta\in\mathcal{T}}{\operatorname{argmax}} \log\Big(L(\theta)\Big) + \log\Big(f_\Theta(\theta)\Big). \tag{4.19b}$$

This is the Maximum a Posteriori (MAP). Compared to the MLE Problem (4.14), the MAP now takes into account the prior as a regularization term.

We now discuss how the estimators (4.14) and (4.19) are related to the *true* value of the model parameters θ_0. To present the statistical properties of the MLE and MAP estimators, let us consider a sequence of N random variables $Y^{(1)}, \ldots, Y^{(N)}$ independently and identically distributed (iid) according to $Y_{\exp}(\chi_{\exp}) = (Y_{\exp}(\chi_1), \ldots, Y_{\exp}(\chi_m))$. It can be shown, under several assumptions, that the MLE estimator is asymptotically consistent and normal. The description of these assumptions can be found in [12, Definition 2.78], [13, Theorems 16.2], and [80, Chapter 7]. One important assumption is that the model parameters θ are identifiable, i.e.,

$$f_{Y|\theta,\chi_{\exp}} = f_{Y|\theta',\chi_{\exp}} \quad \Leftrightarrow \quad \theta = \theta'. \tag{4.20}$$

The identifiability of the model parameters will be revisited in the first numerical example. The book of Cole [81] provides a thorough discussion regarding identifiability issues. The asymptotic consistency of the MLE can be stated as follows: $\forall \varepsilon > 0$,

$$\lim_{N\to\infty} \Pr\left(|\theta_N^* - \theta_0| > \varepsilon\right) = 0. \tag{4.21}$$

This means that the MLE estimator converges in probability to the true value θ_0 of the model parameters. The proofs and precise statement of this result can be found in [13, Theorem 16.2] and [80, 7.1.1]. Moreover, the MLE estimator is asymptotic normal, meaning that the distribution of the random variable $\sqrt{N}(\theta_N^* - \theta_0)$ converges to the normal distribution $\mathcal{N}(0, \mathcal{I}^{-1}(\theta_0))$. Here, the matrix $\mathcal{I}$ refers to the Fisher Information matrix [12, Section 2.3]

$$\mathcal{I}(\theta) = \operatorname{Cov}\Big(\nabla_\theta \ln f_{Y|\theta,\chi_{\exp}}(Y)\Big) = -\,\mathbb{E}\left(H_\theta \ln f_{Y|\theta,\chi_{\exp}}(Y)\right), \tag{4.22}$$

where H_θ denotes the Hessian matrix with respect to θ. An important aspect of the Fisher Information matrix is that it depends upon the experimental design $\chi_{\exp}$. Again the precise statement of the asymptotic normality can be found in [12, Theorem 7.63], [13, Theorem

16.3], [80, Theorem 7.3.1]. Both the asymptotic consistency and normality justify, in parts, the use of the MLE estimator. As the number of data increases, the MLE estimator should converge to the *true* value θ_0 of the model parameters and the variance of the estimator should decrease.

The Bernstein-von-Mises theorem describes the asymptotic properties of the posterior and MAP estimator. Versions of this theorem can be found in [12, Section 7.4], in [13, Chapter 20], and in [56, Chapter 10]. In a nutshell, it can be proved that the MAP estimator is asymptotically consistent and that the conditional distribution of the random variable $\sqrt{n}(\Theta - \theta_0)$ converges to the distribution $\mathcal{N}(0, \mathcal{I}^{-1}(\theta_0))$. These asymptotic results express the fact that the influence of the prior vanishes as the number of data increases.

4.3.2 Calibration in the Presence of Modeling Errors

The calibration problems presented in the previous subsection assumed that the model was perfect. We now alleviate this assumption by assuming that the model may be imperfect, i.e., the model is misspecified and introduces modeling errors. It is important to note that there is no longer a *true* value θ_0 associated with the model parameters. Hence, there is no value of model parameters such that Y^θ matches Y_{exp}. Before proceeding with the presentation of the MLE and Bayesian approach under the KOH framework, we deem useful to revisit the MLE and MAP estimators under misspecified models.

The asymptotic properties of the MLE estimator in the case of a misspecified model have been studied by White [14]. To present these asymptotic results, we need to define the value of the model parameters $\hat{\theta}$

$$\hat{\theta} = \underset{\theta \in \mathcal{T}}{\operatorname{argmin}} \int \ln \left(\frac{f_{Y_{\text{exp}}}(y)}{f_{Y|\theta,\chi_{\text{exp}}}(y)} \right) f_{Y_{\text{exp}}}(y) \, \mathrm{d}y, \tag{4.23}$$

where the objective function is the Kullback-Leibler divergence between the random variables $Y_{\text{exp}}(\chi_{\text{exp}})$ and $Y^\theta(\chi_{\text{exp}})$, noted $\mathrm{KL}(Y_{\text{exp}}(\chi_{\text{exp}}), Y^\theta(\chi_{\text{exp}}))$. White shows [14, Theorem 2.2] that the MLE estimator is an asymptotic estimator of $\hat{\theta}$. Moreover, he also proves [14, Theorem 3.2] that the random variable $\sqrt{N}(\theta_N^* - \hat{\theta})$ converges to $\mathcal{N}(0, C^{-1}(\hat{\theta}))$, for C a variant of the Fisher Information matrix. Similar results have been demonstrated for the Bayesian analysis in [15]. The MAP is a consistent estimator of $\hat{\theta}$, while the conditional distribution of the random variable $\sqrt{N}(\Theta - \hat{\theta})$ converges to a zero-mean normal distribution with covariance matrix akin to that of the Fisher Information matrix.

These two results are important, because they show that even if a model is not perfect, the MLE and MAP estimators will tend to a well defined value. This value $\hat{\theta}$ minimizes the

Kullback-Leibler divergence between the true distribution Y_{exp} and the one obtained with the statistical model Y^{θ}. However, we do not have an explicit representation of the modeling errors by employing directly the estimator (4.14) and (4.19).

Since we shall adopt the KOH framework to explicitly account for modeling errors, through the modeling discrepancy E, we revisit here the MLE and MAP procedures under the KOH relation (4.8). Again, the rationale of employing the KOH approach is that adding the modeling discrepancy E within the formulation yields $Y^{\theta,\xi} \approx Y_{\text{exp}}$. The stochastic process $Y^{\theta,\xi}$ depends upon the model parameters θ and the modeling discrepancy parameters ξ used to describe the modeling discrepancy E^{ξ}. These two types of parameters need to be calibrated, whether with the MLE approach or the Bayesian method. Within the KOH framework, the likelihood function reads

$$L(\theta, \xi) = \prod_{i=1}^{N} f_{Y \,|\, \theta,\xi,\chi_{\text{exp},i}}(y_{\text{exp},i}). \tag{4.24}$$

The maximum likelihood estimation problem now seeks to simultaneously identify the model parameters θ and the modeling discrepancy parameters ξ by solving the following problem, referred here as the MLE KOH problem,

$$(\theta^*, \xi^*) \in \underset{(\theta,\xi)\in\mathcal{T}\times\mathcal{E}}{\text{argmax}}\; L(\theta, \xi). \tag{4.25}$$

Under the assumption that the modeling and observation errors E and Φ are modeled as independent Gaussian processes, the likelihood function in the MLE KOH problem (4.25) can be defined as

$$L(\theta, \xi) = \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^{m_i}\left|\tilde{\Sigma}_i(\xi_k)\right|}} \exp\left(-\frac{1}{2}\|y_{\text{exp},i} - h(\chi_{\text{exp},i}, \theta) - m(\chi_{\text{exp},i}; \xi_m)\|^2_{\tilde{\Sigma}_i(\xi_k)^{-1}}\right), \tag{4.26}$$

where $\tilde{\Sigma}_i(\xi_k) = \Sigma_{E,i}(\xi_k) + \Sigma_{\Phi,i}$. The matrices $\Sigma_{E,i}(\xi_k)$ are the covariance matrix stemming from the Gaussian process E^{ξ} obtained in the same fashion as in (4.15). The maximum likelihood solution for (θ, ξ) is thus computed by solving the minimization problem

$$(\theta^*, \xi^*) \in \underset{(\theta,\xi)\in\mathcal{T}\times\mathcal{E}}{\text{argmin}}\; \frac{1}{2}\sum_{i=1}^{N} \log(\left|\tilde{\Sigma}_i(\xi_k)\right|) + \|y_{\text{exp},i} - h(\chi_{\text{exp},i}, \theta) - m(\chi_{\text{exp},i}; \xi_m)\|^2_{\tilde{\Sigma}_i(\xi_k)^{-1}}. \tag{4.27}$$

We observe that the maximum likelihood solution (θ^*, ξ^*) seeks to minimize the augmented residuals $y_{\text{exp},i} - h(\mathcal{X}_i, \theta) - m(\chi_{\text{exp},i}, \xi_m)$ and the determinant of the resulting covariance matrices $\Sigma_{E,i}(\xi_k) + \Sigma_{\Phi,i}$. The minimization of the augmented residuals is performed by simul-

taneously modifying the model h via the model parameters θ and the mean of the modeling discrepancy via the parameters ξ_m. The presence of the log-determinant of the covariance matrices $\Sigma_{E,i}(\xi_k) + \Sigma_{\Phi,i}$ is important because it prevents some eigenvalues of the covariance matrices $\Sigma_{E,i}(\xi_k)$ to grow too large. Indeed, increasing the eigenvalues of $\Sigma_{E,i}(\xi_k)$ will reduce the importance of the augmented residuals while augmenting the log-determinants. The MLE solution can be seen as the solution that balances the two terms.

The Bayesian approach assimilates the modeling discrepancy parameters ξ as random parameters Ξ. The prior information regarding both the model parameters Θ and modeling discrepancy parameters Ξ is now encoded in the definition of the prior density $f_{\Theta,\Xi}$. The posterior density for (Θ, Ξ) then reads

$$f_{\Theta,\Xi|Y=y_{\text{exp}},\chi_{\text{exp}}}(\theta,\xi) = \frac{L(\theta,\xi)\, f_{\Theta,\Xi}(\theta,\xi)}{\int L(\theta',\xi')\, f_{\Theta,\Xi}(\theta',\xi')\, \mathrm{d}\theta'\, \mathrm{d}\xi'}. \tag{4.28}$$

The goal is often to find the most probable value of the model parameters θ and parameters ξ given the experimental data so that we are interested in the maximum a posteriori problem

$$(\theta^*,\xi^*) \in \underset{(\theta,\xi)\in\mathcal{T}\times\mathcal{E}}{\text{argmax}}\ f_{\Theta,\Xi|Y=y_{\text{exp}},\chi_{\text{exp}}}(\theta,\xi), \tag{4.29a}$$

$$\Leftrightarrow \qquad (\theta^*,\xi^*) \in \underset{(\theta,\xi)\in\mathcal{T}\times\mathcal{E}}{\text{argmax}}\ \log(L(\theta,\xi)) + \log(f_{\Theta,\Xi}(\theta,\xi)). \tag{4.29b}$$

Again, the MAP solution is a regularized version of the MLE, where the prior information is employed as a regularization term. It is usually assumed that the prior information about these two types of parameters is independent, leading to $f_{\Theta,\Xi}(\theta,\xi) = f_\Theta(\theta) f_\Xi(\xi)$. We usually lack prior information about the parameters Ξ, since they relate to the parametrization of the modeling discrepancy, which is a priori unknown. Uninformative or improper priors are typically employed for f_Ξ [18, 21].

Correcting our model observables h with the modeling discrepancy E may introduce some identifiability issues. More precisely, the model parameters θ and the modeling discrepancy parameters ξ may be non-identifiable, meaning that the MLE or MAP for (θ, ξ) is not unique and may belong to a manifold of $\mathcal{T} \times \mathcal{E}$, see e.g. [19, 21, 96]. We will implicitly address this issue using our proposed methodology, as demonstrated in the first example 4.5.1.

4.3.3 Calibration Objective vs Predictive Objective

The following discussion will put in perspective the predictive objective of our model with the aforementioned calibration methods. In the highly idealized case that the model can

represent exactly the system of interest, the calibration process should seek the true value θ_0 of the model parameters. This way, the model is an exact representation of the reality we wish to model, so it can be employed to perform predictions of QoIs at prediction scenarios.

Naturally, no model is perfect, so it is not clear how we should tailor the calibration process toward the predictive goal. As discussed previously, the MLE and MAP estimators converge to the value that minimizes the Kullback-Leibler divergence $\mathrm{KL}(Y_{\mathrm{exp}}(\chi_{\mathrm{exp}}), Y^{\theta}(\chi_{\mathrm{exp}}))$. This is also true in the KOH setting: The MLE estimator (4.25) and MAP estimator (4.29) converge to another value $\hat{\theta}$ that minimizes the Kullback-Leibler divergence between $Y_{\mathrm{exp}}(\chi_{\mathrm{exp}})$ and $Y^{\theta,\xi}(\chi_{\mathrm{exp}})$, pending that θ and ξ are identifiable. However, one can reasonably expect that the divergence $\mathrm{KL}(Y_{\mathrm{exp}}(\chi_{\mathrm{exp}}), Y^{\hat{\theta},\hat{\xi}}(\chi_{\mathrm{exp}}))$ be smaller than $\mathrm{KL}(Y_{\mathrm{exp}}(\chi_{\mathrm{exp}}), Y^{\hat{\theta}}(\chi_{\mathrm{exp}}))$, since the KOH setting enriches the statistical model with the Gaussian process E^{ξ}.

Minimizing the Kullback-Leibler divergence is merely a specific choice of definition for the calibration process. A subtle but important aspect of the Kullback-Leibler divergences $\mathrm{KL}(Y_{\mathrm{exp}}(\chi_{\mathrm{exp}}), Y^{\theta}(\chi_{\mathrm{exp}}))$ and $\mathrm{KL}(Y_{\mathrm{exp}}(\chi_{\mathrm{exp}}), Y^{\theta,\xi}(\chi_{\mathrm{exp}}))$ is that they depend upon the experimental design. Another choice of experimental design χ_{exp} could have led to another optimal value $\hat{\theta}$. Both the choice of the Kullback-Leibler divergence as the notion of dissimilarity and the choice of the experimental design χ_{exp} are not directly related to the predictive objective of the model. In other words, it is far from clear that calibrating the model parameters $\hat{\theta}$ generates a model that allows one to predict an accurate QoI at a prediction scenario.

This assessment could also be translated to the realistic case where we get hold of finite experimental data. The likelihood function is defined by the experimental design χ_{exp} and by the observation noise $\Phi(\chi_{\mathrm{exp}})$. They both define how the residuals are weighted. More precisely, in the case where the observation noise is heteroscedastic, more importance is attributed to fitting the experimental data with small observation noise. Moreover, if the experimental design χ_{exp} samples more densely a region of $\mathcal{X}$, more importance will be put into reducing the residuals in this region. With the MAP, the introduction of a prior over Θ (and potentially Ξ) regularizes in a certain way the calibration solution, by weighting directly these parameters. The rationales behind these weighting processes and regularizations defining the calibration functional are based upon statistical and sampling considerations, not upon the predictive objective of the model. It will be beneficial to have a greater control over the objective function of the calibration problem, so that we can better relate it to the prediction objective.

4.4 Goal-Oriented Calibration

In this Section, we define our goal-oriented approach to the calibration problem. A key feature of our methodology consists in clearly separating the calibration of the model parameters θ and the calibration of the modeling discrepancy parameters ξ. This separation stems from the fact that the nature and the purpose of these two types of parameters are significantly different. On one hand, the calibration of the modeling discrepancy E must reflect its empirical nature. Given a specific model (with given model parameters θ), the objective of the modeling discrepancy E is to adjust the relation (4.8) so that the experimental data y_{exp} may reasonably come from $Y^{\theta,\xi}$. The calibration of the modeling discrepancy E is then viewed as a regression problem. This regression problem must take into account statistical assumptions regarding the sampling of the experimental data. On the other hand, the calibration of the model parameters θ must be tailored toward the predictive objective of the model. The parameters ξ describing the modeling discrepancy E are thus assimilated as parameters to be optimized for each value of the model parameters θ. In the next two sections, we will first describe two approaches for the calibration of the modeling discrepancy parameters ξ given a value of the model parameters θ. Then we will describe the goal-oriented calibration problem.

4.4.1 Calibration of the Modeling Discrepancy

As mentioned earlier, we adopt the point of view that for each value of model parameter θ, there is an optimal representation of the modeling discrepancy E based on the KOH formulation (4.8). For a given representation of the modeling discrepancy E, we need to calibrate the parameters ξ. We consider here two methods for the calibration of the modeling discrepancy E. The first one consists in solving the Maximum Likelihood problem

$$\xi^*(\theta) = \Big(\xi_m^*(\theta), \xi_k^*(\theta)\Big) \in \underset{(\xi_m,\xi_k)\in\mathcal{E}_m\times\mathcal{E}_k}{\operatorname{argmin}} L\Big(\theta, (\xi_m, \xi_k)\Big). \tag{4.30}$$

This approach is akin to the one proposed by Leoni et al. [21]. By solving Problem (4.30), we simultaneously find the parameters ξ_m^* and ξ_k^* of the modeling discrepancy that make the observations y_{exp} the most probable. It is important to note that the optimal parameters ξ^* are obtained for a given value of the model parameters θ. We note that this Maximum Likelihood Estimation problem must be solved for each value of θ.

The second approach to calibrate the modeling discrepancy is akin to the first approach, with the important modification that we calibrate ξ_m and ξ_k in a sequential manner. More precisely, we first calibrate the mean function m via its parameters ξ_m by assuming that

there is no covariance term. The KOH relation (4.8) can thus be simplified as

$$Y^{\theta,\xi_m}(\chi) = h(\chi,\theta) + m(\chi;\xi_m) + \Phi(\chi), \tag{4.31}$$

so as to generate a likelihood function $L_m : \mathcal{T}\times\mathcal{E}_m \to \mathbb{R}$. We then calibrate the parameters ξ_m via the Maximum Likelihood Estimation

$$\xi_m^*(\theta) \in \underset{\xi_m\in\mathcal{E}_m}{\operatorname{argmax}}\, L_m(\theta,\xi_m). \tag{4.32}$$

This procedure can be viewed as calibrating the first moment of the Gaussian process E, while assuming that its second moment vanishes. We then calibrate the second moment of the modeling discrepancy E, namely the covariance function k, by assuming the now calibrated mean $m(\cdot;\xi_m^*(\theta))$. To this end, we assume that the modeling discrepancy is distributed as $E^{\xi_m^*(\theta),\xi_k} \sim \mathcal{GP}(m(\chi;\xi_m^*(\theta)), k(\chi,\chi';\xi_k))$. The KOH relation now reads

$$Y^{\theta,\xi_k}(\chi) = h(\chi,\theta) + E^{\xi_m^*(\theta),\xi_k}(\chi) + \Phi(\chi), \tag{4.33}$$

which in turn provides the new likelihood function $L_k : \mathcal{T}\times\mathcal{E}_k \to \mathbb{R}$. The parameters ξ_k are then calibrated in terms of the Maximum Likelihood problem

$$\xi_k^*(\theta) \in \underset{\xi_k\in\mathcal{E}_k}{\operatorname{argmax}}\, L_k(\theta,\xi_k) \tag{4.34}$$

The proposed procedure can be summarized as follows: we initially calibrate the first moment of the Gaussian process E, then we employ the resulted mean to calibrate its second moment. It is important to emphasize again that the modeling discrepancy parameters $\xi^*(\theta) = (\xi_m^*(\theta),\xi_k^*(\theta))$ are calibrated for a given value of the model parameters θ.

These two approaches for the calibration of the modeling discrepancy E both employ MLE problems to optimize the parameters ξ. However, the first approach simultaneously calibrates ξ_m and ξ_k, while the second approach calibrates them in a sequential manner. Both procedures generate optimal modeling discrepancy parameters $\xi^*(\theta)$, so a final KOH relation can be written as

$$Y^{\theta}(\chi) = h(\chi,\theta) + E^{\xi^*(\theta)}(\chi) + \Phi(\chi), \tag{4.35}$$

where $E^{\xi^*(\theta)} \sim \mathcal{GP}(m(\chi;\xi_m^*(\theta)), k(\chi,\chi';\xi_k^*(\theta)))$. Hence, the augmented model prediction Y^θ no longer depends on the value of the modeling discrepancy parameters ξ.

We mention that one could also have chosen the Bayesian framework to calibrate the mod-

eling discrepancy parameters ξ. However, since the parameters ξ pertain to the modeling discrepancy that is a priori unknown, assigning an informative prior f_Ξ is far from trivial. This is the reason why we adopt here the MLE framework. We now analyze in greater detail the Maximum Likelihood problems (4.30), (4.32), and (4.34). An important point to emphasize here is that these optimization problems are solved for a fixed value of model parameters θ. Thus, there is no need to solve/call the model, usually a costly procedure. In the case where the mean and the kernel of the Gaussian process E depend smoothly upon the parameters ξ, then the loglikelihood also depends smoothly upon ξ. The optimization problems (4.32) and (4.34) can then be efficiently solved using a derivative-based solver. However, the objective function may possess multiple local minima. Then, if one employs a local solver, multiple starting points should be considered to ensure that a good local minimum is reached. For this study, we use the solver Knitro [72], which implements an interior-point algorithm. Moreover, this solver possesses the built-in capability to perform multi-start optimization.

4.4.2 Definition of the Goal-Oriented Calibration Problem

We now address the problem of calibrating the model parameters θ via a goal-oriented approach. The idea consists in directly using the (calibrated) modeling discrepancy $E^{\xi^*(\theta)}$ to define a calibration objective functional J for the model parameters θ. The calibration problem of the model parameters θ reads in this case

$$\theta^* \in \underset{\theta \in \mathcal{T}}{\operatorname{argmin}}\, J\left(E^{\xi^*(\theta)}\right). \tag{4.36}$$

One option to define the functional J is to exploit the fact that the random variable $E^{\xi^*(\theta)}(\chi)$ follows a normal law with mean $m(\chi;\xi_m^*)$ and variance $k(\chi,\chi;\xi_k^*)$. For each parameter χ, we can define the quantile function as

$$Q_{E(\chi)}(\beta) = F^{-1}_{E(\chi)}(\beta), \tag{4.37}$$

where $F_{E(\chi)}$ represents the cumulative distribution function of the normal random variable $E(\chi)$ and β the quantile. The interval $[Q_{E(\chi)}((1-\alpha)/2), Q_{E(\chi)}((1+\alpha)/2)]$ specifies the α-prediction interval, i.e., the error will fall at a frequency α within this interval. We can then seek the probable maximal absolute error at level α, which describes our notion of error

$$\varepsilon_E(\chi,\alpha) = \max\left\{\left|Q_{E(\chi)}((1-\alpha)/2)\right|, \left|Q_{E(\chi)}((1+\alpha)/2)\right|\right\}. \tag{4.38}$$

The function (4.38) is referred here to as the *quantile error function*. A visual representation of this function for $\alpha = 0.95$ is shown in Figure 4.2.

We then define the calibration functional J, for a given level α, as

$$J\left(E^{\xi^*(\theta)}\right) = \|\varepsilon_{E^{\xi^*(\theta)}}(\cdot, \alpha)\|. \tag{4.39}$$

The introduction of this objective functional J allows us to directly relate the predictive objective of the model to the calibration process via the choice of the norm. We now list some predictive objectives with their associated calibration objective:

- we suppose that the model is to be employed to make predictions over a set $\mathcal{X}_{\text{pred}}$. The calibration process should then be tailored toward minimizing the modeling discrepancy over $\mathcal{X}_{\text{pred}}$. The norm could then be chosen as the $L^2(\mathcal{X}_{\text{pred}})$ norm;
- we suppose that the model is to be employed to make predictions at a specific parameter χ_{pred}. The calibration process should then be tailored toward minimizing the modeling discrepancy around this specific parameter χ_{pred}. The norm could then be chosen as a weighted norm, such as $\|g\| = \int_{\mathcal{X}} w(\chi) g^2(\chi)\, \mathrm{d}\chi$, where w is a weighting function.

The introduction of this calibration functional is a crucial departure from the calibration methods described in Section 4.3. As mentioned earlier, the weights attributed to the modeling discrepancy with the MLE and Bayesian frameworks are based upon statistical considerations and experimental designs. With the proposed goal-oriented approach, the statistical assumptions and the experimental design are still taken into account in the calibration of the modeling discrepancy E (via problems (4.30), (4.32), and (4.34)). However, we now have control over the weights we put on to the modeling discrepancy when calibrating the model parameters θ. Again, because the model parameters θ and the modeling discrepancy parameters ξ are different in nature, we adopt separate calibration approaches. It is important to emphasize that the previous formulations of the objective functions for the goal-oriented calibration are but possible definitions. The important point of all these previous calibration formulations is that they are designed to minimize some particular features of the modeling discrepancy E.

To solve Problem (4.36), we will utilize the blackbox optimization software NOMAD [74]. For each value of θ, we need to calibrate the modeling discrepancy E with dedicated optimization problems.

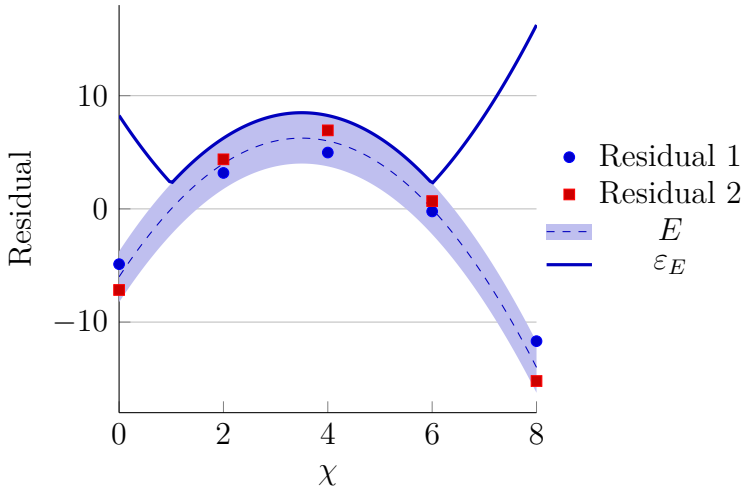


Figure 4.2 Illustration of the quantile error function ε_E from the modeling discrepancy E (illustrated in blue). The dashed curve indicates the mean function of the Gaussian process while the shaded area represent the 95% prediction interval. Two sets of residuals are displayed for illustrative purpose.

4.5 Numerical Examples

We demonstrate our calibration methodology on two synthetic examples and on an application problem. The first example considers a simple model where experimental data are generated from a synthetic model. This example will be employed to study the various aspects of our goal-oriented approach. In particular, we will focus on:

1. the impact of the method employed for the calibration of the modeling discrepancy;
2. the definition of the calibration goals;
3. the impact of the representation of the modeling discrepancy;
4. the comparison of the goal-oriented calibration when using Problems (4.14) and (4.25).

The second synthetic problem consists in a verification example. We will verify that, in the ideal case where the model is perfect (absence of modeling errors), our goal-oriented approach is able to capture the "true" model parameters θ. Finally, we will consider the thermal problem of the Sandia Validation Challenge mentioned throughout the previous section and will illustrate how a goal-oriented calibration can be tailored toward several predictive objectives.

4.5.1 Synthetic Example

We revisit the problem presented by Brynjarsdóttir and O'Hagan [19] as our synthetic problem. The example consists in a simple linear model

$$h(\chi, \theta) = \theta x, \tag{4.40}$$

for $\chi = x$, $x \in [0, 6]$, and $\theta \in [0.3, 0.7]$. The above model is a surrogate model used to describe the system of interest, whose true behavior is provided by the synthetic model

$$h_{\text{synt}}(\chi) = \frac{0.65x}{1 + x/10}, \tag{4.41}$$

where the curvature have been slightly increased compared to [19]. The experimental data are corrupted by the experimental noise $\Phi \sim \mathcal{GP}(0, 0.01^2\delta(\chi, \chi'))$ (zero-mean white noise of variance 0.01^2). We acquire data for ten equidistant values of x and perform three replicates of the experiment ($\chi_{\text{exp},1} = \cdots = \chi_{\text{exp},3} = (0, 2/3, \ldots, 16/3, 18/3)$). We consider that the (synthetic) data within a replicate are correlated, but we remind the reader that the data for different replicates are independent. Figure 4.3 shows the experimental data as well the model prediction obtained with $\theta = 0.45$. The sparsity of the data and the low number of replicates are chosen so as to reflect a practical situation.

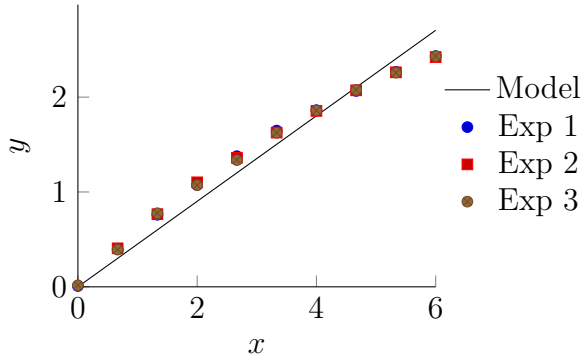


Figure 4.3 Experimental data and model prediction for $\theta = 0.45$.

Calibration of the Modeling discrepancy

We study here the calibration of the modeling discrepancy parameters ξ via the two approaches described in Section 4.4.1. In order to illustrate these calibration approaches, we represent the modeling discrepancy E with the Gaussian process $E^\xi \sim \mathcal{GP}(m(\chi; \xi_m), k(\chi, \chi'; \xi_k))$,

where

$$m(\chi;\xi_m) = \xi_{m,1} + \xi_{m,2}x, \tag{4.42a}$$

$$k(\chi,\chi';\xi_k) = \sigma(\chi;\xi_k)\sigma(\chi';\xi_k)\exp\left(-\frac{1}{2}\frac{(x-x')^2}{\xi_{k,1}^2}\right), \tag{4.42b}$$

with $\sigma(\chi;\xi_k) = \log(1+\exp(\xi_{k,2}+\xi_{k,3}x))$. This construction of the modeling discrepancy is inspired by [70, 71]. It allows the specification of a non-stationary Gaussian process via function σ. This function is a smooth approximation of the ReLU function whose purpose is to ensure that the kernel remains positive definite. A stationary process can be recovered by setting $\xi_{k,3} = 0$. For this application, we decide to model the length scale $\xi_{k,1}$ as a constant value. Using this representation, we can model a linear mean and a quasi-linear variance for the modeling discrepancy.

With the aforementioned choice of experimental noise, we solve Problem (4.30) and Problems (4.32) and (4.34) for $\theta = 0.449$ and for $\mathcal{E}_m = [-50, 50] \times [-50, 50]$ and $\mathcal{E}_k = [-10, 10] \times [-10, 100] \times [-10, 100]$. Figure 4.4 illustrates the optimal value of ξ obtained with the two approaches.

From Figure 4.4, we observe that the two approaches for calibrating the modeling discrepancy E yield different results. For the method calibrating simultaneously ξ_m and ξ_k via Problem (4.30), we note that the mean of the Gaussian process does not pass through the residual. Moreover, it increases significantly the variance in order to cover the residual. This phenomenon is accompanied by a notable increase in the length scale $\xi_{k,1}$ (up to its upper bound), as shown by the realization of the Gaussian process. This somewhat peculiar result can be explained by recalling Equation (4.26). The MLE solution will try to increase as much as possible the variance of the Gaussian process in order to minimize the importance of the data misfit. This increase is however balanced by the logdeterminant of the resulting covariance matrix. For this specific application, it happens that the optimal solution is the one shown in Figure 4.4. On the other hand, the sequential approach fits the first moment of the Gaussian process via the MLE Problem (4.32), producing a mean function that passes through the residual. Then, the sequential method fits the second moment based on the previously computed mean. The variance and the length scale are thus computed based on this constraint.

The value of the loglikelihood with the first method is 69.7 and 50.2 for the sequential approach. The density function of the observables thus takes higher value with the first approach than with the second. Based solely upon this consideration, the first approach to calibrate the modeling discrepancy should be favored over the second. Indeed, the sequential

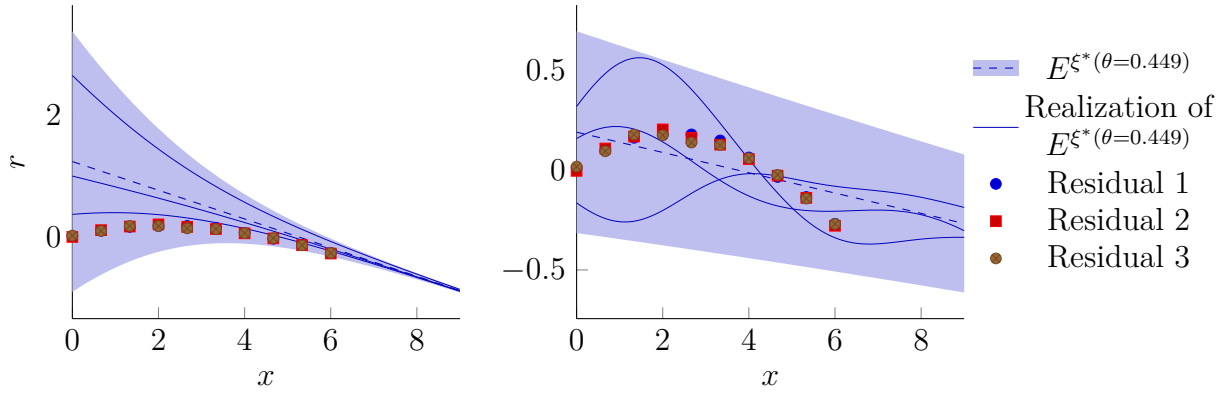


Figure 4.4 (Left) Optimal representation of the modeling discrepancy E obtained by solving Problem (4.30). (Right) Optimal representation of the modeling discrepancy E obtained by sequentially solving problems (4.32) and (4.34). Both representation of the modeling discrepancy are obtained for $\theta = 0.449$. The dashed blue line is the mean of the Gaussian process while the blue-shaded area is the 95% prediction interval. Some realizations of the priors are displayed by thin lines.

approach can be viewed as a constrained version of the first method, where the mean function is constrained to better reflect the first moment of the data. However, our goal-oriented approach for the calibration of the model parameters θ directly relies on the quantile error function (4.38). Figure 4.5 illustrates the function (4.38) for the two representations of the modeling discrepancy. We observe that the value of $\varepsilon_E(0, 0.95)$ is larger for the first approach than with the second. This is somewhat problematic considering that the residuals at $x = 0$ are very small and tightly spread. Nevertheless, we believe that it is more important to obtain a representation of the modeling discrepancy E that avoids increasing the variance of the modeling discrepancy E. We will then consider only the sequential approach to calibrate the modeling discrepancy E from now on. We also mention here that it is important to constrain the range of the length scale with respect to the experimental design. For example, a lower bound for the length scale should reflect the minimal distance between the experimental data. This prevents the possibility of obtaining modeling discrepancy that features artificially high frequencies. In the same manner, an upper bound for the length scale should also be defined as to reflect the information provided by the experimental data. Again, this mitigates the possibility of obtaining modeling discrepancy with artificially low frequencies. One can view these bound constraints as specifying a prior on Ξ, via the definition of $\mathcal{E}_k$.

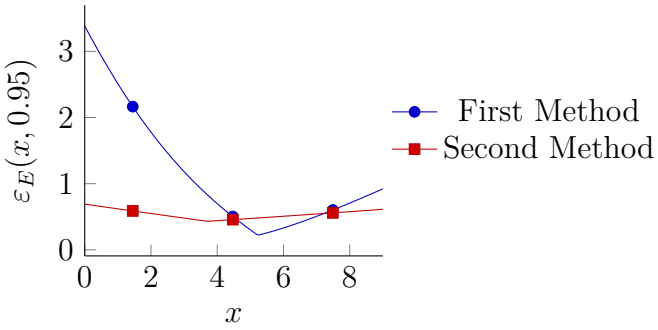


Figure 4.5 Quantile error function (4.38) for the modeling discrepancy obtained with the first and second calibration approaches. The value of α is fixed at 95%.

Goal-Oriented Calibration

Now that the subproblem of calibrating the parameters ξ of the modeling discrepancy E has been illustrated, we can focus on solving the goal-oriented calibration problem. We test the goal-oriented calibration approach for the three predictive objectives described in Table 4.2.

The first predictive objective states that we want to employ the model to perform predictions for $x \in [0, 6]$. It is then reasonable to define the calibration goal as the minimization of the modeling discrepancy E over $x \in [0, 6]$. The second predictive objective stipulates that the model will be employed to make predictions near $x = 0$. The calibration goal could then be defined as minimizing a weighted functional of the modeling discrepancy, where greater emphasis is put on x near the origin. Finally, the last predictive objective consists in employing the model to perform predictions for which no experimental data is gathered. This predictive objective can be viewed as an extrapolation. For each value of θ considered, the modeling discrepancy $E^{\xi^*(\theta)}$ is extrapolated up to $x = 9$. The third calibration goal is defined as the minimization of the weighted modeling discrepancy, where greater importance is given to the extrapolated modeling discrepancy near $x = 9$. We compute the three objective

Table 4.2 Choice of the norm for the functional J

i	Predictive Objective	Norm
1	The model will be employed to perform predictions over $x \in [0, 6]$	$\|g\|^2 = \int_0^6 g(x)^2 \, \mathrm{d}x$
2	The model will be employed to perform predictions near $x = 0$	$\|g\|^2 = \int_0^6 \exp\left(-\frac{1}{2}\frac{x^2}{0.75^2}\right) g(x)^2 \, \mathrm{d}x$
3	The model will be employed to perform predictions near $x = 9$ (Extrapolation)	$\|g\|^2 = \int_0^9 \exp\left(-\frac{1}{2}\frac{(x-9)^2}{0.75^2}\right) g(x)^2 \, \mathrm{d}x$

functions for $\theta \in [0.3, 0.7]$. The profiles are displayed in Figure 4.6. It is important to note that the particular scaling of the weight functions presented in Table 4.2 does not impact the calibration process, since it consists in mere scaling of the objective function. We observe that the optimal values for the model parameters θ vary with the definition of the calibration functional. These results echo our discussion regarding model parameters θ: since no model is perfect, there are no true values of model parameters θ, only optimal ones with respect to a given goal. Different notions of optimality will lead to different optimal model parameters θ^*. Using the goal-oriented approach, we explicitly define the notion of optimality by minimizing some weighted functionals of the modeling discrepancy E. The optimal value of the model parameters θ are $\theta \approx 0.445$, $\theta \approx 0.611$, and $\theta = 0.421$ for the first, second, and third calibration objective, respectively.

We further analyze the goal-oriented approach by presenting three optimal solutions alongside their corresponding modeling discrepancy E and quantile error functions (4.38). These results are shown in Figure 4.7. For the first objective, we note that the optimal model does indeed attempt to minimize the mean error for $x \in [0, 6]$. With the second objective, the optimal model does minimize the modeling discrepancy near $x = 0$. Finally, the optimal model obtained with the last objective slightly decrease the value of θ compared to the optimal model obtained with the first calibration objective. Given the linear mean of the modeling discrepancy E as its kernel (a quasi-linear variance), minimizing the modeling discrepancy at $x = 9$ is performed by minimizing the modeling discrepancy near $x = 6$. A side-by-side comparison of the modeling discrepancies and their associated quantile error functions (4.38) provide a visual interpretation of the calibration results. Each optimal model attempts to minimize a specific feature of the quantile error function. On one hand, the optimal model for the second objective is the one possessing a smaller quantile error function near $x = 0$ compared to the other two optimal solutions. On the other hand, the third optimal solution indeed minimizes the quantile error function for large x. Finally, the first model with respect to the first objective provides the smallest quantile error function in the L^2 norm.

Impact of the Representation of the Modeling Discrepancy

A key aspect of the goal-oriented calibration relies on the calibration of the modeling discrepancy. We verify here the influence of the choice of representation of the modeling discrepancy on the whole goal-oriented procedure. To this end, we consider this additional representation

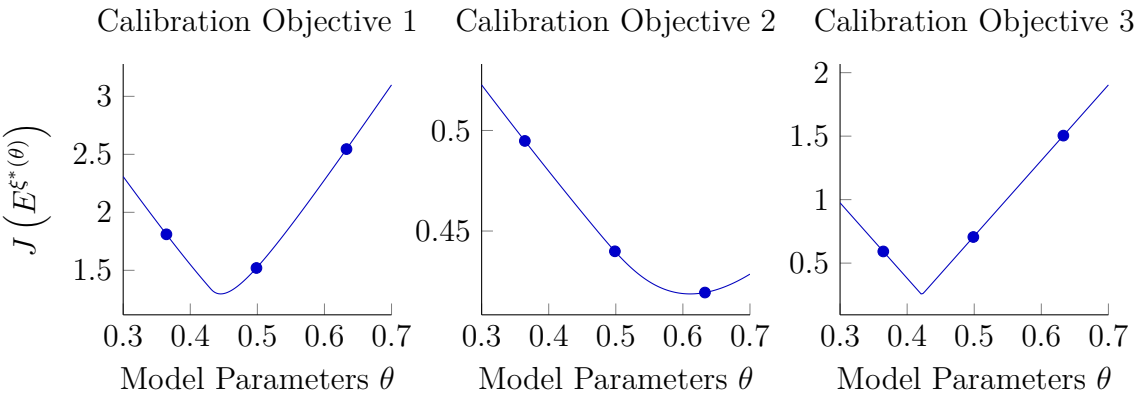


Figure 4.6 Profiles of the calibration functional for the first objective (left), the second objective (middle), and the third objective (right).

of the modeling discrepancy $\tilde{E}^{\tilde{\xi}} \sim \mathcal{GP}(\tilde{m}(\chi;\tilde{\xi}_m), \tilde{k}(\chi,\chi';\tilde{\xi}_k))$:

$$\tilde{m}(\chi;\tilde{\xi}_m) = \tilde{\xi}_{m,1}, \tag{4.43a}$$

$$\tilde{k}(\chi,\chi';\tilde{\xi}_k) = \tilde{\xi}_{k,2} \exp\left(-\frac{1}{2}\frac{(x-x')^2}{\tilde{\xi}_{k,1}^2}\right). \tag{4.43b}$$

This representation of the modeling discrepancy is simpler than the one presented in (4.42). We can now only capture the constant mean of the modeling discrepancy and the variance is considered constant across χ. The Gaussian process is stationary in this case. With this simpler representation of the modeling discrepancy, we compute the three objective functions considered in Section 4.5.1. The resulting profiles are presented in Figure 4.8.

Several observations can be made from Figure 4.8. First, the three profiles are similar. These results are in contrast with those presented in Figure 4.6. This can be explained by the fact that the representation of the modeling discrepancy is too simple. The mean and variance of the modeling discrepancy E cannot vary with χ, so all optimization problems seek the same optimal model, irrespective of the weight function considered. The optimal model obtained is essentially the one that minimizes the maximal residual. Second, we note that the profiles are discontinuous. This stems from the way the modeling discrepancy is calibrated. As mentioned previously, the MLE problem (4.34) may present multiple local maxima. The discontinuity indicates that the global maximum ξ^* has jumped from one local maximum to another.

These results indicate important requirements for our goal-oriented calibration method. If we want to weight the modeling discrepancy with respect to the parameters χ, it is essential to allow the modeling discrepancy to become non-stationary. This can be performed via the

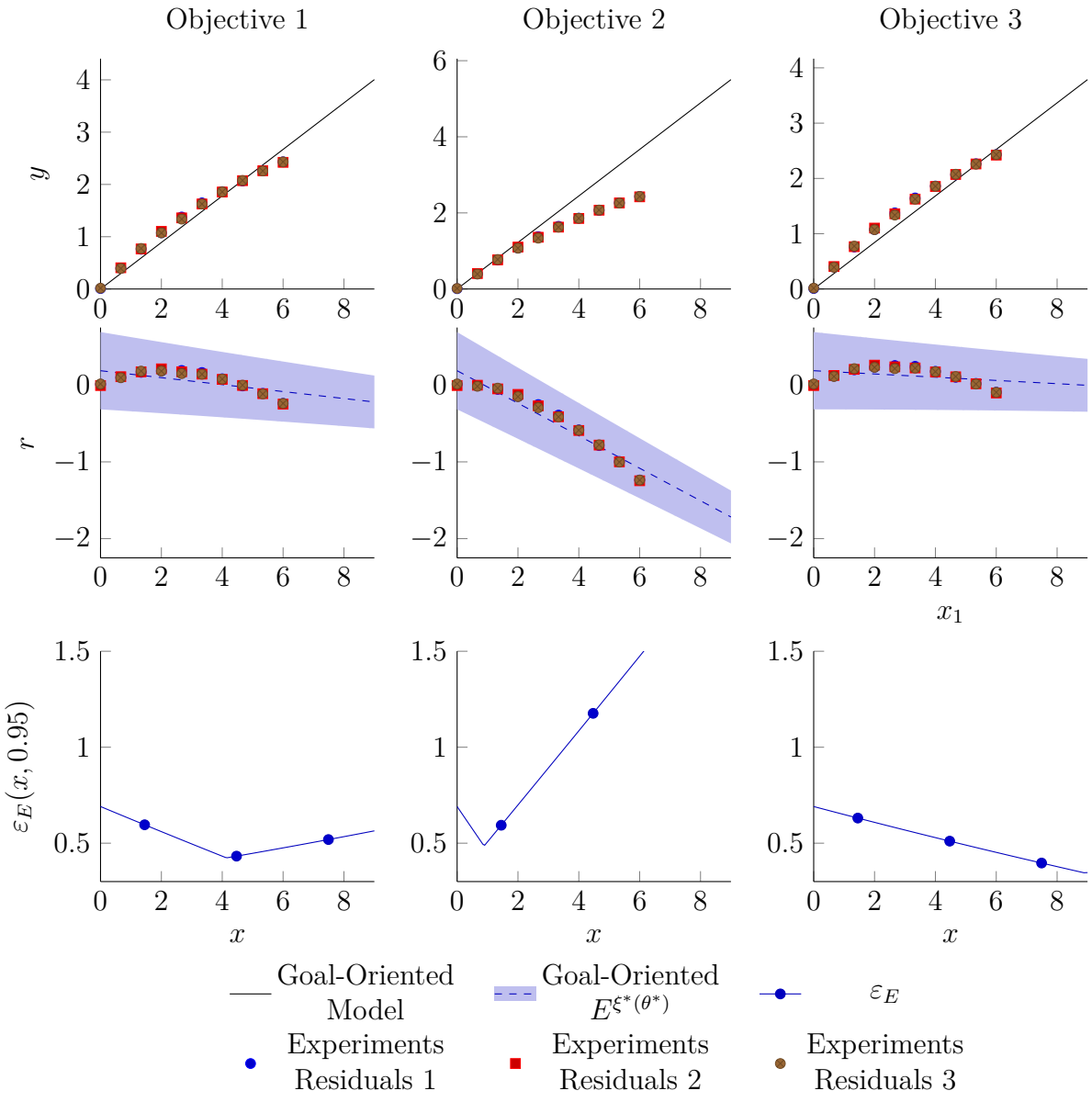


Figure 4.7 Goal-oriented calibration for the first objective (left), the second objective (middle), and the third objective (right). The top row illustrates the optimal model predictions. The middle row presents the associated residuals and modeling discrepancy $E^{\xi^*(\theta^*)}$. The bottom row displays the corresponding quantile error functions.

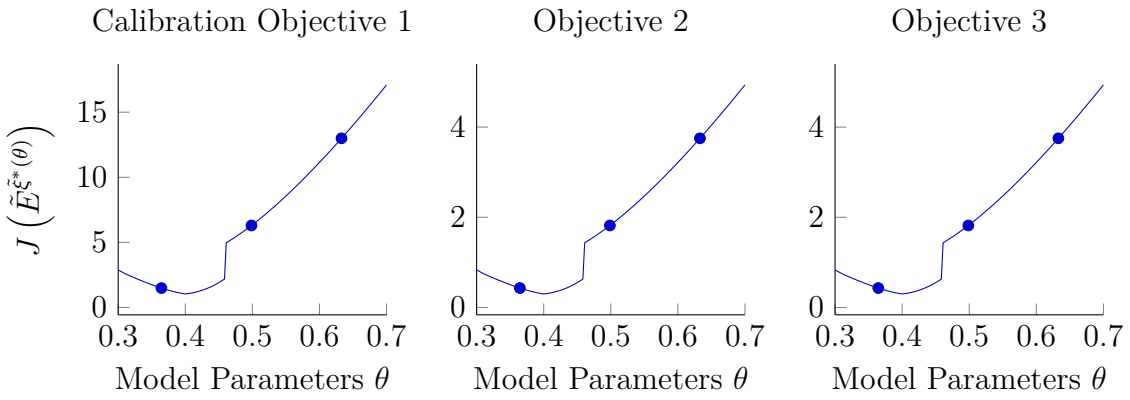


Figure 4.8 Profiles of the calibration functional for the first objective (left), the second objective (middle), and the third objective (right) with the representation (4.43) of the modeling discrepancy.

mean function m and the covariance function k of the modeling discrepancy E. Moreover, since our approach relies on the approximation of the modeling discrepancy, it is important to consider a representation rich enough to capture the important features missed by the model.

Comparison of the Calibration Methods

We compare the previous calibration results with the ubiquitous MLE problem (4.14). The optimal value of model parameters θ with the maximization of the likelihood (4.13) as our definition of optimality is $\theta = 0.444$. We note that this optimal value is very close to the optimal value obtained with the first goal-oriented calibration. These results are coherent, since the MLE problem in this particular case reduces to a least squares problem where equal weight are attributed to all residuals. The two calibration objectives are then similar.

We also solve the MLE problem (4.14) for two curated sets of experimental data. The first set is only comprised of experimental data gathered at $x \in \{0, 2/3, 4/3\}$. This ad hoc selection is performed so that only experimental data in the regime near $x = 0$ are considered. With this new definition of optimality, we obtain $\theta^* = 0.580$. This slope is slightly lower than that obtained with the goal-oriented calibration for the second predictive objective ($\theta^* \approx 0.611$). The second set of experimental data only considers data gathered at $x \in \{14/3, 16/3, 6\}$. They are selected because they are closest to $x = 9$. The optimal value for the model parameters is $\theta^* = 0.421$, which happens to be very similar to the optimal value obtained with the third goal-oriented calibration. These results highlight the fact that different notions of optimality can lead to similar optimal values of the model parameters.

These MLE solutions rely on ad hoc selection of experimental data to specify their notion of optimality. On the other hand, the proposed goal-oriented calibration method does consider all the experimental data gathered to calibrate the modeling discrepancy E. The notion of optimality is also transparently described in the goal-oriented calibration method.

We now solve the MLE KOH problem (4.25) with the modeling discrepancy defined in (4.42). We encounter two issues when solving this particular calibration problem. First, a clear identifiability issue [81] appears when calibrating simultaneously the model parameters θ and the modeling discrepancy parameters ξ with the MLE. Indeed, both the model parameter θ and the parameter $\xi_{m,2}$ consist in linear terms with respect to parameter x. The MLE problem calibrates the sum $\theta + \xi_{m,2}$ to the optimal value $\theta + \xi_{m,2} = 0.214$. There is not a unique global maximum (θ^*, ξ^*) for the MLE KOH problem (4.25). Thus, the MLE KOH solutions can not distinguish the individual contribution of the model observable functional h and the modeling discrepancy E. The model parameters θ and the modeling discrepancy parameters ξ are then considered non-identifiable. If the predictive objective is to employ the model to perform predictions on the observable (with the functional h), the identifiability issue is not a problem *per se*. For any value of the model parameters θ among the optimal ones, the observable predictions can always be corrected by the corresponding optimal representation of the modeling discrepancy. However, if one wishes to predict a QoI defined such that $h_{\text{qoi}} \neq h$, then the identifiability issue becomes important. Indeed, there are no longer modeling discrepancy to correct the QoI. In this sense, two different optimal values for the model parameters would yield the same observable predictions if corrected by their respective modeling discrepancy, but they could yield two very different QoI predictions. Without additional knowledge on which model parameters to choose among the optimal ones, there is no clear way to decide which QoI predictions should be employed.

The second issue is reminiscent of the one raised in Section 4.5.1 with the calibration of the modeling discrepancy when simultaneously considering the mean and kernel parameters. In fact, concurrently calibrating the model and the modeling discrepancy amounts to calibrate a sole Gaussian process with mean function $h(\chi, \theta) + m(\chi; \xi_m)$ and kernel function $k(\chi, \chi'; \xi_k)$. The same phenomenon also shows up when the optimal modeling discrepancy parameters ξ consist of a large length scale and a large variance. Let us consider for example the possible optimal model parameters $\theta^* = 0.3$ and the associated optimal parameters $\xi^*_{m,2} = -0.086$. The corresponding model predictions and modeling discrepancies are displayed in Figure 4.9. From the left plot of Figure 4.9, we observe that the Gaussian process Y_{exp} does encompass the experimental data. However, the mean function of Y_{exp} is far from the actual experimental data and we note that the variance is very large near $x = 0$. Here, the model prediction (black line) lays far from the experimental data. The right plot shows the residuals and

the modeling discrepancy Gaussian process $E^{\xi^*(\theta=0.3)}$. The modeling discrepancy shows the same behavior as on the left plot of Figure 4.4, where the variance is increased and the mean function does not fit the residuals. It is important to reiterate here that MLE KOH problem (4.25) in this example yields non-identifiable model parameters θ and modeling discrepancy parameters ξ. Thus, the calibration results presented in Figure 4.9 lead to but one possible optimal solution. Due to the identifiability issue, we can conclude that the MLE KOH calibration problem (4.25) is not well-posed for this specific example.

4.5.2 Verification Example

In the highly idealized case where the model is able to perfectly capture the response of the system of interest, the calibration process should yield the true model parameters. Again, we reiterate our point of view that for any practical physical system of interest, modeling errors will always be present. The notion of true model parameters should then be replaced by the notion of optimal model parameters. However, with the following synthetic example, we can investigate if our goal-oriented calibration of the model parameters θ is able to recover the true model parameters. Since our approach allows some freedom in the definition of the calibration objective, we also need to verify if, whatever our reasonable definition of optimality, we recover the true model parameters.

The proposed model features two parameters $\chi = (x_1, x_2) \in \mathcal{X} = [0, 2] \times [0, 10]$ and four model parameters $\theta = (\theta_1, \ldots, \theta_4) \in \mathcal{T} = [0, 2]^4$. The observation functional is defined as

$$h(\chi, \theta) = \exp(-\theta_1 x_2) \left(0.5 \sin(x_1 - \theta_2 x_2) + \theta_3 \exp(-0.5(x_1 - \theta_4)^2)\right). \tag{4.44}$$

To create the synthetic data, we consider the true model parameters $\theta_0 = (0.15, 0.8, 1.2, 0.25)$. The design of experiments is a lattice consisting of five linearly-spaced data in the x_1 direction and six linearly-spaced data in the x_2 direction. We corrupt the data with observational noise $\Phi \sim \mathcal{GP}(0, 0.01^2 \delta(\chi, \chi'))$. The noise is chosen small as to mitigate its impact on the calibration process. The dataset includes two independent replicates. Figure 4.10 illustrates the response surface (4.44) as well as the experimental design.

The representation of the modeling discrepancy is given by $E^{\xi} \sim \mathcal{GP}(m(\chi; \xi_m), k(\chi, \chi'; \xi_k))$

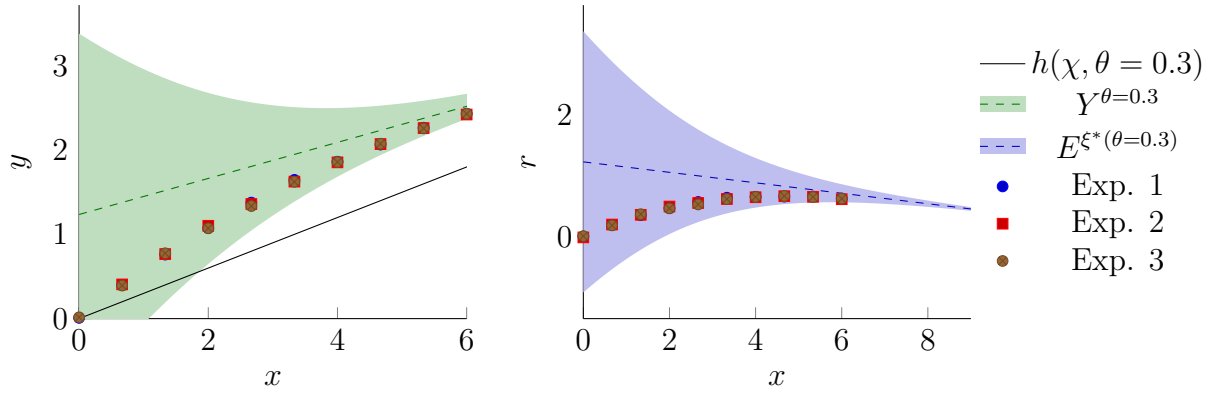


Figure 4.9 One possible optimal solution to the calibration Problem (4.25). The optimal model prediction and associated modeling discrepancy are displayed for the optimal model parameters $\theta^* = 0.3$.

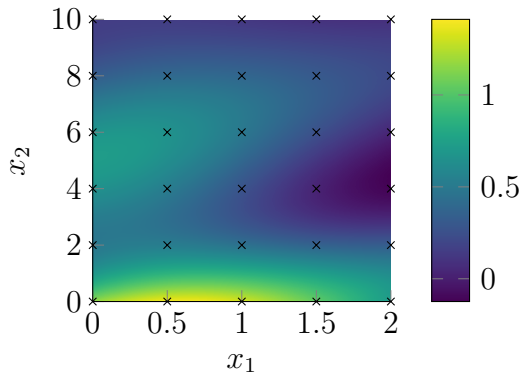


Figure 4.10 Surface of response of the manufactured solution and the design of experiments marked by black crosses.

with:

$$m(\chi; \xi_m) = \xi_{m,1} + \xi_{m,2} x_1 + \xi_{m,3} x_2, \tag{4.45a}$$

$$k(\chi, \chi'; \xi_k) = \sigma(\chi)\sigma(\chi') \exp\left(-\frac{1}{2} \frac{\|\chi - \chi'\|_2^2}{\xi_{k,1}^2}\right), \tag{4.45b}$$

$$\sigma(\chi) = \log(1 + \exp(\xi_{k,2} + \xi_{k,3} x_1 + \xi_{k,4} x_2)). \tag{4.45c}$$

The goal is to verify our calibration procedure for $\alpha = 0.95$ and for three calibration objectives defined for three predictive objectives, as described in Table 4.3. We thus solve Problem (4.36) with the various calibration objectives. For the sake of completeness, we also solve the MLE problem (4.14) and MLE KOH problem (4.25). We compile in Table 4.4 the optimal model parameters θ^* and in Table 4.5 the modeling discrepancy parameters $\xi^*(\theta^*)$ that were

obtained. We observe that the relative errors for the three optimal model parameters θ^* obtained with the goal-oriented calibration are relatively small. This indicates that the proposed goal-oriented calibration is able to recover the true model parameters θ_0. If one chooses a denser experimental design and more replicates, one would expect the relative error on the model parameters to vanish. The optimal parameters $\xi_m^*(\theta^*)$ show that the mean of the modeling discrepancy is very small. Moreover, the components of the parameters $\xi_k^*(\theta^*)$ determining the variances (the second, third, and fourth components) indicate that the spread of the modeling discrepancy E is small. The value of the optimal length scales $\xi_{k,1}^*(\theta^*)$ is of no particular importance here. Since both the mean and the variance are small, we may conclude that the goal-oriented approach has indeed recognized that there is no modeling discrepancy in this specific application. We note that these results hold for the three objectives, even the third one that attempted to minimize an extrapolation of the modeling discrepancy. However, it is important to mention that the solution of the optimization Problem (4.36) for the third objective necessitates more tuning of the solver NOMAD, hinting that the problem is more difficult to solve.

The MLE problem (4.14) and the MLE KOH problem (4.25) are also able to recover the true model parameters θ_0, as their respective relative errors are small. The optimal parameters ξ^* computed by solving the MLE problem (4.25) also yield vanishing modeling discrepancy E. It is interesting to observe that in the idealized case of no modeling errors, all calibration problems are able to recover the true model parameters θ_0. Hence, we may conjecture that if the modeling errors are negligible, the precise definition of the calibration objective does not impact significantly the calibration results. However, if the modeling errors are non-negligible, then the definition of the calibration problem has a direct influence on the calibrated model parameters. Again, any model describing a physical system will generate some modeling errors. We seldom know a priori if the modeling errors are small, thus it is important to ponder the choice of the calibration method and its associated notion of optimality.

Table 4.3 Choice of the norm for the functional J

i	Predictive Objective	Norm
1	The model will be employed to perform predictions over $\mathcal{X}$	$\|g\|^2 = \int_{\mathcal{X}} g(\chi)^2 \, \mathrm{d}\chi$
2	The model will be employed to perform predictions near $\tilde{\chi} = (1.3, 4)$	$\|g\|^2 = \int_{\mathcal{X}} \exp\left(-\frac{1}{2}\frac{\|\chi-\tilde{\chi}\|_2^2}{0.3^2}\right) g(\chi)^2 \, \mathrm{d}\chi$
3	The model will be employed to extrapolate predictions near $\tilde{\chi} = (2.5, 12)$	$\|g\|^2 = \int_{[0,2.5]\times[0,12]} \exp\left(-\frac{1}{2}\frac{\|\chi-\tilde{\chi}\|_2^2}{0.75^2}\right) g(\chi)^2 \, \mathrm{d}\chi$

Table 4.4 Optimal model parameters θ^* for different calibration objectives

	θ^*	$\|\theta_0 - \theta^*\|_2 / \|\theta_0\|_2$
Obj. 1	(0.151, 0.797, 1.21, 0.235)	1.24%
Obj. 2	(0.151, 0.800, 1.21, 0.234)	1.28%
Obj. 3	(0.151, 0.793, 1.21, 0.227)	1.77%
MLE (4.14)	(0.151, 0.798, 1.21, 0.242)	0.659%
MLE KOH (4.25)	(0.153, 0.799, 1.19, 0.258)	0.862%

Table 4.5 Optimal modeling discrepancy parameters $\xi^*(\theta^*)$ for different calibration objectives

	$\xi_m^*(\theta^*)$	$\xi_k^*(\theta^*)$
Obj. 1	(-1.19e-3, -1.62e-4, 1.05e-4)	(2.23e-2, -4.17, -5.34, -5.25)
Obj. 2	(-1.67e-3, 1.85e-4, 2.18e-4)	(10.0, -4.19, -10.0, -1.01)
Obj. 3	(8.30e-4, 1.20e-3, -3.21e-4)	(10.0, -3.93, -10.0, -1.16)
MLE (4.14)	-	-
MLE KOH (4.25)	(1.38e-2, -1.01e-2, 1.26e-4)	(10.0, -5.01, -10.0, -0.249)

4.5.3 Sandia Validation Challenge - Thermal Problem

We now consider the application described in Section 4.2, namely the thermal problem presented at the Sandia Validation Challenge 2006 [1, 42]. With this example, we illustrate how the calibration objective can impact the predictive capabilities. Our objective here is not to provide an answer to the Sandia Validation Challenge. We will merely exploit the given synthetic data. We have chosen this application as a representative of a real case-study. Indeed, the experimental data are gathered at different conditions than that of a scenario of interest. Moreover, two sets of experimental data are provided. The first one (noted ensemble validation in [1]) will be employed to perform the calibration of the model parameters $\theta = (k, \rho C_p)$. The second set (noted accreditation test) is obtained at experimental data closer to the scenario of interest. This second set of experimental data will be employed to assess the predictive capabilities of the calibrated model. It is important to emphasize that our study does not seek to alter or improve the model (4.1). In this sense, we assume that the thermal conductivity k and the volumetric heat capacity ρC_p are scalar parameters to be calibrated.

We complement the experimental design mentioned in Section 4.2.3 with the additional data given in [1]. More precisely, the experimental design for the calibration data now consists of 16 independent experiments. Each experiment in the calibration set consists of 11 data gathered at equidistant times instances $t \in [0, 1000]$. The experimental design

is described in Table 4.6. We note that experimental data are available for two different values of heat flux q and length L and multiple values of time t. Hence, we can capture a linear mean in the modeling discrepancy E in these directions. For the same reason, we allow the variance to vary quasi-linearly with respect to q, L, and t. Within an independent experiment, we assume that the kernel function k depends only on time t. We suspect that the modeling discrepancy E should behave differently at $t = 0$ from its behavior at larger t, because the initial temperature can be accurately represented by the model. Hence, we allow the length-scale of the Gaussian kernel to vary linearly with respect to the parameter t. With these considerations, we suppose that the modeling discrepancy can be modeled as $E^\xi \sim \mathcal{GP}(m(\chi;\xi_m), k(\chi,\chi';\xi_k))$ where

$$m(\chi;\xi_m) = \xi_{m,1} + \xi_{m,2}q + \xi_{m,3}L + \xi_{m,4}t,$$

$$k(\chi,\chi';\xi_k) = \sigma(\chi;\xi_k)\,\sigma(\chi';\xi_k)\,\ell(\chi';\xi_k)^{\frac{1}{4}}\,\ell(\chi';\xi_k)^{\frac{1}{4}}\left(\frac{\ell(\chi,\xi_k)+\ell(\chi',\xi_k)}{2}\right)^{-\frac{1}{2}}\exp\left(-q(\chi,\chi';\xi_k)\right),$$

$$\text{with } \sigma(\chi;\xi_k) = \log\left(1+\exp\left(\xi_{k,3}+\xi_{k,4}q+\xi_{k,5}L+\xi_{k,6}t\right)\right),$$

$$\ell(\chi,\xi_k) = \log\left(1+\exp\left(\xi_{k,1}+\xi_{k,2}t\right)\right),$$

$$q(\chi,\chi',\xi_k) = \left(\frac{\ell(\chi,\xi_k)+\ell(\chi',\xi_k)}{2}\right)^{-1}(t-t')^2.$$

Again, the construction of such a Gaussian process is inspired from [70, 71]. We note the presence of normalizing terms so that the standard deviation of the Gaussian process is directly described by the function σ.

Two experiments are conducted at $\chi_{17,j}$ and $\chi_{18,j}$, which consist in accreditation data. The corresponding experimental data are collected in Table 8 of [1]. They are coined accreditation data since they are obtained at conditions closer to that of the prediction scenario. As mentioned in Section 4.2.3, we assume that the observation error is $\Phi \sim \mathcal{GP}(0, 0.01^2\delta(\chi,\chi'))$.

Table 4.6 Experimental design for the calibration and accreditation data

i	$\chi_{i,j}$	Use
$1,\ldots,4$	$(1000, 0.0127, 25, 0, t)$ for $t \in \{0, 100, \ldots, 1000\}$	Calibration
$5,\ldots,8$	$(1000, 0.0254, 25, 0, t)$ for $t \in \{0, 100, \ldots, 1000\}$	Calibration
$9,\ldots,12$	$(2000, 0.0127, 25, 0, t)$ for $t \in \{0, 100, \ldots, 1000\}$	Calibration
$13,\ldots,16$	$(2000, 0.0254, 25, 0, t)$ for $t \in \{0, 100, \ldots, 1000\}$	Calibration
$17,\ldots,18$	$(3000, 0.019, 25, x, t)$ for $t \in \{0, 50, 100, \ldots, 1000\}$ and for $x \in \{0, 0.0095, 0.019\}$	Accreditation

We study below the goal-oriented calibration under two distinct predictive settings.

Prediction over the Calibration Set

For the first predictive setting, we wish to employ the model to perform predictions for parameters $(q, L, u_0, x, t) \in \mathcal{X}_{\text{calib}} = [1000, 2000] \times [0.0127, 0.0254] \times \{25\} \times \{0\} \times [0, 1000]$. It is important to note that the 16 independent experiments were performed for parameters in $\mathcal{X}_{\text{calib}}$. Hence, we use an interpolation of the modeling discrepancy for the goal-oriented calibration. We translate this predictive objective into the following norm for the quantile error function

$$\|g\| = \int_{1000}^{2000} \int_{0.0127}^{0.0254} \int_{0}^{1000} g(q, L, 25, 0, t)^2 \, \mathrm{d}t \, \mathrm{d}L \, \mathrm{d}q. \tag{4.46}$$

We solve the goal-oriented problem (4.36) alongside the MLE problem (4.14) and MLE KOH problem (4.25).

We shall compare the calibration results according to three criteria. The first is the likelihood of the data given the assumption of each calibration approach. The second criterion is the root mean-squared error $\text{RMSE}_{\text{calib}}$ computed with the experimental data used for the calibration, i.e.,

$$\text{RMSE}_{\text{calib}} = \sqrt{\sum_{i=1}^{16} \sum_{j=1}^{11} \left(h(\chi_{i,j}, \theta^*) - y_{i,j}\right)^2}. \tag{4.47}$$

The third criterion is the norm (4.46) of the quantile error function. For the MLE approach without modeling discrepancy (4.14), we reconstruct a posteriori the modeling discrepancy. This is done according to the sequential approach described in Section 4.4.1 for the optimal value of the computed model parameters θ^*. The values of the criteria are displayed in Table 4.7. We observe that the calibration approach yielding the smallest RMSE is the MLE problem without modeling discrepancy (4.14). This result is expected, since, with the assumption of homoscedastic observational noise, this calibration problem attempts to minimize directly the mean-square error (MSE). The goal-oriented calibrated model gives only a slightly larger RMSE. Since its objective is to minimize the modeling discrepancy over $\mathcal{X}_{\text{calib}}$, the goal-oriented approach indirectly seeks to lower this RMSE. On the other hand, the MLE KOH approach gives a large RMSE. This calibration approach does not discriminate the model and the modeling discrepancy. It simply attempts to maximize the likelihood by calibrating the combined model parameters and modeling discrepancy.

We observe that the MLE KOH solution is the one producing the largest loglikelihood. Both

Table 4.7 $\text{RMSE}_{\text{calib}}$, loglikelihood, and functional of the quantile error function for the calibration approaches

	MLE	MLE KOH	Objective 1
Optimal Model Parameters θ^*	(0.0547, 4.06e5)	(0.0477, 3.96e5)	(0.0516, 4.22e5)
$\text{RMSE}_{\text{calib}}$	**13.9**	25.6	14.1
Loglikelihood	-1.71e8	**-187**	-220
$J\left(E^{\xi^*(\theta^*)}\right)/1e4$	37.2*	54.8	**33.9**

the MLE and MLE KOH problems maximize the loglikelihood, but the latter enriches the representation of Y with the modeling discrepancy E, via relation (4.8). It is then expected that with a richer representation, the optimal loglikelihood will be increased. Also, the loglikelihood for the MLE KOH calibration is several orders greater than that with the MLE approach, indicating that the experimental data are significantly more likely to be explained by the KOH representation (4.8) than by the representation (4.12), which takes into account only the model.

The goal-oriented solution is the one minimizing the quantile error function over $\mathcal{X}_{\text{calib}}$. This result had to be expected, since it is precisely this functional that the goal-oriented approach seeks to reduce. We note that with the modeling discrepancy $E^{\xi^*(\theta)}$ constructed a posteriori for θ^* computed with the MLE method, we obtain a slightly larger value of $J\left(E^{\xi^*(\theta^*)}\right)$. Since it minimizes the RMSE, the a posteriori calibration of the modeling discrepancy captures the fact that the residuals are small at the experimental design parameters. According to the minimization of the quantile error function, the MLE KOH provides the worst calibrated model. Again, this stems from the fact that this calibration method does not differentiate the model from the modeling discrepancy, while the quantile error function depends only upon the modeling discrepancy. The key takeaway from these results is that the choice of the calibration method induces its own notion of optimality. Each model parameter shown in Table 4.7 is optimal with respect to its definition of optimality.

Prediction at the Prediction Scenario

The final objective described in the Sandia Thermal Challenge [1, Section 1] consists in predicting the temperature at the prediction scenarios $\chi_{\text{pred}} = (3500, 0.019, 25, x, 1000)$, for $x \in [0, 0.019]$. As one can check, the accreditation scenarios are indeed closer to the parameter space $\mathcal{X}$ than the calibration scenarios. The accreditation data will be employed

here to assess the quality of the calibrated model with respect to the predictive objective. To this end, we will compute the RMSE with respect to the accreditation data

$$\mathrm{RMSE}_{\mathrm{accred}} = \sqrt{\sum_{i=17}^{18}\sum_{j=1}^{63}\left(h(\chi_{i,j},\theta^*) - y_{i,j}\right)^2}. \tag{4.48}$$

It is important to remind the reader that the accreditation data are not used in any way in the calibration process. They are only used to assess the quality of the calibrated models. In order to tailor the goal-oriented calibration toward the second predictive objective, we choose the following norm for the quantile error function

$$\|g\| = \int_{1000}^{3500}\int_{0.0127}^{0.0254}\int_{0}^{1000} w(q,L,t)\, g(q,L,25,0,t)^2 \,\mathrm{d}t\,\mathrm{d}L\,\mathrm{d}q, \tag{4.49}$$

where

$$w(q,L,t) = \exp\left(-\frac{1}{2}\left[\frac{(q-3500)^2}{200^2} + \frac{(L-0.019)^2}{0.002^2} + \frac{(t-1000)^2}{200^2}\right]\right). \tag{4.50}$$

The norm puts more weight on the modeling discrepancy near $(q, L, t) = (3500, 0.019, 1000)$. We employ here an extrapolation of the modeling discrepancy: for each value of θ, we calibrate the parameters of the modeling discrepancy to obtain $\xi^*(\theta)$ and we extrapolate this error to the region of interest. We calibrate the model parameters θ according to the goal-oriented problem (4.39) for $\alpha = 0.95$ and the aforementioned norm of the quantile error function. Figure 4.11 shows the model prediction and the modeling discrepancy at the four configurations at which the data are gathered.

We observe that even if the calibration objective is to reduce the quantile error function near the prediction scenario, the Gaussian process $Y^{\xi^*(\theta^*)}$ still covers well the calibration experimental data. This illustrates the central aspect of the goal-oriented methodology: for each value of θ, we find an optimal representation of the modeling discrepancy. Then we evaluate the norm of the corresponding quantile error function. The optimal model parameters θ^* are obtained by minimizing the extrapolation of the quantile error function. By comparing the prediction obtained using the model calibrated with MLE, we observe that it indeed minimizes the MSE. The results are different from those for the model calibrated using the goal-oriented approach with the second objective, but the differences are rather small. The predictions of the model calibrated with the MLE KOH exhibit a relatively poor fitting of the data. Again, this is due to the fact that the method simultaneously calibrates the model parameters and the modeling discrepancy without discrimination. The modeling discrepancy $E^{\xi^*(\theta^*)}$ obtained with the goal-oriented approach for the second objective do

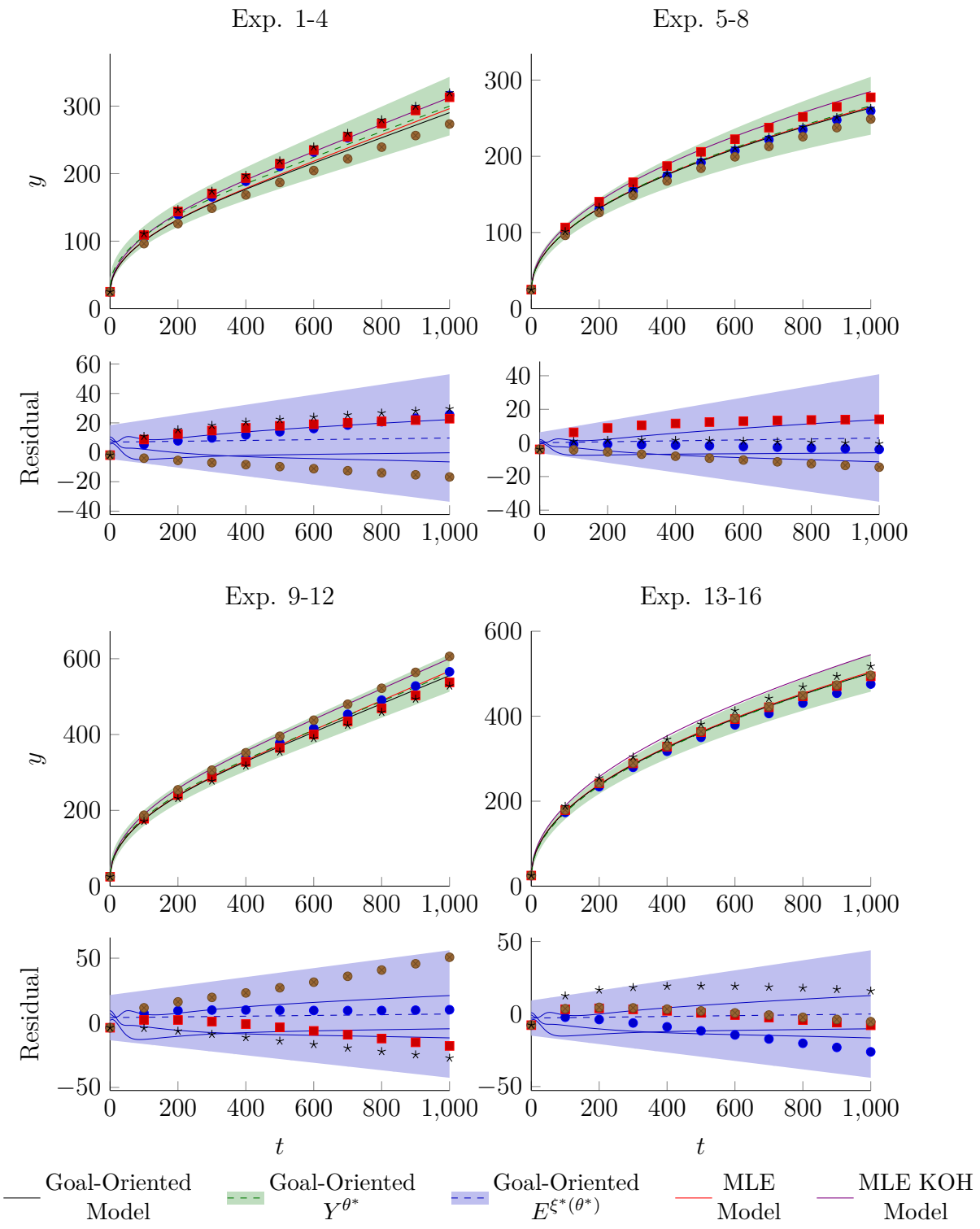


Figure 4.11 Optimal model predictions and associated modeling discrepancy at the four calibration scenarios. The first and third rows show the predictions obtained with the various calibration methods. The second and fourth rows depict the modeling discrepancies and residuals obtained with the second objective of the goal-oriented calibration. Some realizations of the modeling discrepancies $E^{\xi^*(\theta^*)}$ are also demonstrated by thin curves.

encompass the residuals. The mean of the modeling discrepancy captures relatively well the linear behavior with respect to the parameters (q, L, t). The variance of the modeling discrepancy $E^{\xi^*(\theta^*)}$ increases with time t. This result does reflect the behavior of the residuals. Finally, we observe that the length-scale of the Gaussian kernel also increases with time t. On one hand, the residuals vary significantly between $t = 0$ and $t = 100$, so the length-scale is shorter there. On the other hand, the residuals do not deviate as much for larger time t, thus explaining the larger length-scale in this region.

We now compare the calibrated models against the accreditation data. It is important to reiterate that none of these calibration approaches have used the accreditation data to compute their respective optimal model parameters θ^*. Figure 4.12 shows the accreditation data with the calibrated models obtained with MLE, MLE KOH, and the goal-oriented approach with the second objective. We note that the stochastic prediction Y^{θ^*} obtained with the goal-oriented formulation nearly covers the accreditation data. Importantly, the mean of the modeling discrepancy is of small magnitude, since the model predictions (in black) and the mean of the stochastic prediction Y^{θ^*} (in green) are nearly superposed, especially for large values of t. This is explained by the fact that we specifically designed the goal-oriented calibration to reduce the quantile error function at the prediction scenario, which is near the accreditation scenario. The prediction confidence, depicted in shaded green, is not vanishingly small, because the modeling discrepancy still needs to fit the calibration data, as depicted in Figure 4.11.

By comparing the $\text{RMSE}_{\text{accred}}$ displayed in Table 4.8, we conclude that the MLE calibrated model performs better at the accreditation scenario. This can also be observed by looking at the red curves in Figure 4.12. Again, the MLE calibration problem does not directly seek to minimize the residuals at the accreditation scenario. However, it appears that by reducing the RMSE at the calibration scenarios, it also reduces the modeling discrepancy at the accreditation scenario. The MLE KOH problem yields the larger $\text{RMSE}_{\text{accred}}$ due to the large residuals produced at the accreditation scenario for $x = 0$, as seen in Figure 4.12 (left). The second goal-oriented calibrated model is the second best model according to $\text{RMSE}_{\text{accred}}$. It is interesting to note that, even though the calibration tailors the model to reduce the (extrapolated) modeling discrepancy at the prediction scenario, it is still able to reduce the modeling discrepancy at the accreditation scenario. However, the first goal-oriented calibrated model, where we seek to reduce the modeling discrepancy over $\mathcal{X}_{\text{calib}}$, does not capture as well the accreditation data as for the case with the second objective. This result indicates that the behavior of the modeling discrepancy potentially varies in the parameter space $\mathcal{X}$. The MLE calibrated model and the second goal-oriented calibrated model appear to minimize relevant features of the modeling discrepancy for the prediction of the temperature

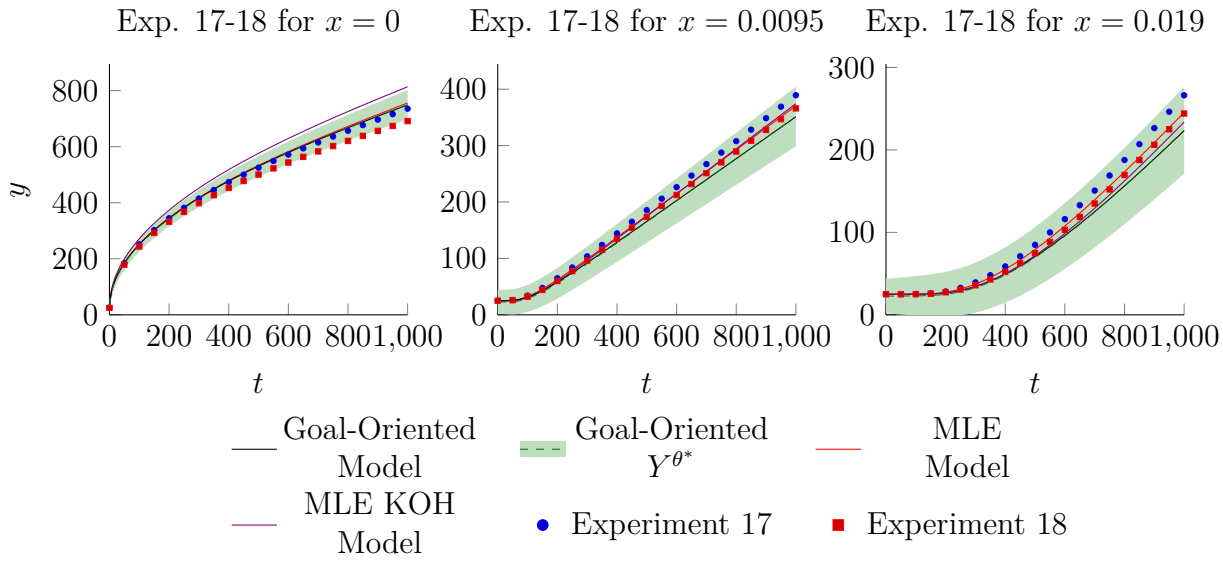


Figure 4.12 Accreditation data and model predictions for the second goal-oriented calibrated model (black), the MLE calibrated model (red) and the MLE KOH calibrated model (violet). The Gaussian process Y^{θ^*} for the second goal-oriented calibration is displayed in green.

at the accreditation scenario.

4.6 Conclusion

In this paper, we have proposed a novel method to the calibration problem. First, we have presented a description of a model and its different types of parameters. We have described various aspects of the process of collecting experimental data, namely the experimental design and the observation noise. We have employed the KOH relation (4.8) to assimilate the experimental data with the model and the modeling discrepancy. We have presented two ubiquitous approaches to the calibration problem, the Maximum Likelihood Estimation and the Bayesian Inference. Both methods are described with and without modeling errors. Then, we have introduced our calibration methodology, which relies on the fact that there is no true

Table 4.8 $\text{RMSE}_{\text{accred}}$ for the calibration approaches

	MLE	MLE KOH	Objective 1	Objective 2
Optimal Model Parameters θ^*	(0.0547, 4.06e5)	(0.0477, 3.96e5)	(0.0516, 4.22e5)	(0.0528, 4.26e5)
$\text{RMSE}_{\text{accred}}$	**17.4**	40.1	22.9	20.1

model parameters θ, but rather optimal ones. The first aspect of our goal-oriented approach consists in calibrating the parameters ξ of the modeling discrepancy E^{ξ} for each value of the model parameters θ. Two approaches have been proposed to handle this task, whether by calibrating simultaneously the first and second moments of the Gaussian process E^{ξ}, or by calibrating them in a sequential manner. The calibration of the model parameters θ is then performed by minimizing a weighted norm of the modeling discrepancy, whose weights can be chosen in terms of the predictive objective of the model.

We have applied the goal-oriented calibration on a synthetic problem to illustrate various properties of the method. Comparison between the two calibration methods of the modeling discrepancy parameters ξ shows that the sequential approach presents some attractive features when compared to the direct approach. Then, we have illustrated the goal-oriented calibration with different calibration objectives. The results show that the optimal model parameters θ^* do indeed directly depend upon the notion of optimality. However, these results are conditional upon a sufficiently rich representation of the modeling discrepancy E. The manufactured example allows the verification of the goal-oriented calibration in the very idealized case free of modeling errors. For the last example, we have applied the goal-oriented calibration approach on the thermal problem of the Sandia Validation Challenge. By comparing the results obtained with the MLE, MLE KOH, and goal-oriented formulations, we have highlighted the fact that we indeed seek optimal model parameters, whose value directly changes with the calibration objective. Finally, we have employed the goal-oriented strategy with a calibration objective specifically tailored toward the minimization of the modeling discrepancy at a prediction scenario. By comparing the predictions computed using the calibrated models with the accreditation data, it appears possible to improve the predictive capabilities of the model by minimizing key features of the modeling discrepancy.

This last observation offers an interesting avenue for future research. With the proposed goal-oriented calibration method, we are now capable of choosing the optimal model parameters that minimize features of the modeling discrepancy. However, the method in itself does not identify the relevant features of the modeling discrepancy that affect the QoI. Indeed, it is not clear how the modeling discrepancy on the observables are propagated to the modeling discrepancy on the QoI. This is especially true if:

- The QoI consists in a prediction at a non-reproducible scenario, e.g. prediction at future time;
- The QoI consists in non-observable quantities, such as maximal/minimal quantities.

A possible way to relate the observables at the calibration scenarios to the QoI at the pre-

diction scenario could be given by adapting the methodology presented in [41]. With this approach, the weighting function could then be directly defined with respect to the prediction of the QoI. Pursuing this avenue could remove the somewhat arbitrary choices made in this paper regarding the form of the weighted norm. This would prove very beneficial if the calibration data are gathered at different scenarios and for different types of observables.

As shown in Example 4.5.1, the representation of the modeling discrepancy may impact the results of the goal-oriented calibration. Further work should explore a more robust way to define the representation of the modeling discrepancy. One possible avenue is to consider a collection of representations that could be compared according to model selection criteria [35, 37, 38]. Along the same line, other ways to calibrate the modeling discrepancy parameters ξ could be investigated. The proposed sequential approach appears to mitigate the issue of jumping from a local maximum to another one, but further work should investigate this behavior.

Related to the previous point is the necessity to further examine whether the proposed goal-oriented calibration defines a well-posed problem. More precisely, for a given functional of the quantile error function, the question is whether an optimal model parameter θ^* exists and is unique. The goal-oriented methodology appears to alleviate the identifiability issue between the model parameters and parameters of the modeling discrepancy. However, it may be that the choice of the functional introduces some identifiability issues that may make the optimal solution θ^* non-unique. Finally, the stability of the solution with respect to the functional of the quantile error and the experimental data should be investigated.

A final avenue for future research could be to modify the modeling assumptions regarding the model parameters. In this work, we considered the model parameters to be deterministic, but one can envision to alleviate this hypothesis by assimilating the model parameters as random variables. The challenge would then consist in combining the KOH formulation with the approach proposed, for example, in [25]. In this case, some variability in the data can be represented by the model while the modeling discrepancy will capture the remaining errors that cannot be solely explained by the structure of the model.

CHAPTER 5 CONCLUSION

In this thesis, we have presented novel goal-oriented formulations of the validation and the calibration processes. In scientific computing, one usually wishes to employ a model to predict some specific QoIs which are not directly observable from the system of interest. Although several works pertaining to the estimation and control of numerical errors in QoIs have been proposed in the past, validation and calibration methodologies have not been extensively studied with respect to the prediction of QoIs. Extending validation processes toward the prediction of QoIs is an important task since the whole modeling process aims at using mathematical models to achieve specific engineering objectives. Even if comparing model predictions with experimental data does provide useful information about the validity of model observables, one needs to make a step further and make sure that the assessment can be extended to evaluate whether QoIs predicted by the same model are indeed accurate. Tailoring the calibration process toward the prediction of QoIs is also important, but in a more subtle way. The calibration process allows the computation of optimal values of the model parameters, which in turn affects the modeling errors produced. We need to make sure that the notion of optimality induced by the calibration approach attempts to reduce the modeling errors affecting the QoIs.

5.1 Summary of Works

The first contribution of this thesis, presented in Chapter 3, focuses on adapting the validation process to the prediction of non-observable QoIs. We recognize that any validation process should compare the model predictions with experimental data. The first novel aspect of our work consists in considering the selection of the experiments as an inherent part of the validation process. The process of designing validation experiments should be performed a priori of any experimental data collection. This particular point is a departure from previous studies that assess the relevance of experimental data a posteriori. The second original aspect proposes a systematic way to design these validation experiments. More precisely, the proposed methodology selects the observables employed for the validation such that they are the most representative of the QoI. To quantify this notion of representative, we introduce influence matrices M_{qoi} and M_{obs} based on the Active Subspace sensitivity method. The proposed approach to design optimal validation experiments for the prediction of QoI is separated in two steps. The first step computes the optimal validation scenario x_{val} by minimizing the distance with the influence matrix of the QoI at the prediction scenario. The

second step computes the optimal observation functional $h_{\text{obs,val}}$ and sensor parameters z_{val} by minimizing the distance between the normalized influence matrices of the observable with that of the QoI at the validation scenario. A synthetic example illustrates the importance of taking into account both the QoI and the prediction scenario in the choice of the observable used for the validation. A tentative validation experiment wrongly yields that the model is likely valid, while the optimal validation experiment allows us to correctly assess that the model is not valid for the prediction purpose. A second example studying the transport of a pollutant in a fictitious river illustrates the importance of the definition of the QoI in the optimal validation experiment to be performed.

Chapter 4 presents the second contribution of this thesis, namely making the calibration process goal-oriented. A central component of our proposed methodology lies on the following line of reasoning: since no model is perfect, it is not well-founded to seek the true value of the model parameters θ. Instead, the calibration process computes an optimal value θ^* for the model parameters based on a notion of optimality implicitly or explicitly defined. By acknowledging that no model is perfect, we implicitly acknowledge that there exists modeling errors associated to the model which therefore must also be accounted for. We propose a new and original view of the modeling discrepancy: for each value of the model parameters θ, we generate a specific model which in turn possesses its own modeling discrepancy. This shift of paradigm leads us to the original proposition of defining the calibration objective as a functional of the approximation of the modeling discrepancy E. In this sense, we deviate from the MLE and from the Bayesian approaches of the calibration problem. The proposed goal-oriented method calibrates, for each value of model parameters θ, the parameters ξ of the modeling discrepancy with MLE approaches. With this calibrated modeling discrepancy $E^{\xi^*(\theta)}$ we then evaluate the calibration functional tailored toward the predictive goal. We illustrate the proposed calibration approach on a simple synthetic problem. We observe that a sequential calibration of the modeling discrepancy, where the first and second moment of the Gaussian process are sequentially computed, presents beneficial features compared to a direct MLE calibration. Moreover, we highlight a key requirement for the goal-oriented approach in that the representation of the modeling discrepancy E must be sufficiently rich and non-stationary. The study of a manufactured problem provides a sanity check in the highly idealized case where the model is a perfect representation of reality (no modeling errors). Finally, we employ the data provided for the thermal problem at the Sandia Validation Challenge to test multiple calibration objectives with the goal-oriented calibration. This example reinforces the point of view that by reducing certain features of the modeling discrepancy on the observables, we may decrease also the modeling discrepancy affecting the QoI.

5.2 Limitations and Perspectives

The proposed goal-oriented calibration methodology allows one to calibrate the model parameters θ such that a functional of the modeling discrepancy E^{ξ^*} is minimized. A significant drawback of the approach is that the definition of the calibration functional still depends on somewhat arbitrary considerations. We may sensibly relate this functional with the prediction goal if it pertains to the prediction of observables, as it is done in the numerical examples. But still, some arbitrary decisions were made, especially on how the weighting functions vary in the parameter space. Moreover, if the QoI is non-observable, then it is not obvious how to define the calibration functional toward this predictive goal. An interesting avenue for future research consists in applying the comparison of the behavior of the QoI with that of the observable for the goal-oriented calibration. More precisely, one can envision defining the weighting function of the calibration functional by computing the distance between the influence matrix of the QoI at the prediction scenario with the influence matrices of the observables. In this way, one could also ponder for which type of observables the modeling discrepancy should be reduced.

This work on goal-oriented calibration also highlights the importance of the approximation of the modeling discrepancy E. Further research should be devoted to the sub-problem dealing with the calibration of the modeling discrepancy. For example, the sequential approach could further be extended by sequentially calibrating the parameters defining the mean, variance, and covariance terms of the Gaussian process E. Moreover, it is not clear a priori which representation of the modeling discrepancy is rich enough considering the experimental design and the experimental data at hand. Multiple representations could be compared by ways of model selection using the Akaike Information Criteria or the Bayes factors.

The idea of computing influence matrices to compare the behavior of QoIs and observables at various scenarios ought to be applied in other settings. For example, one could contemplate the analysis, for a given model, of existing experimental data via their influence matrices. It should then be interesting to employ this information to extend cross-validation approaches so that they are tailored toward the prediction of QoI.

We believe that these methodologies are sufficiently versatile to be used in real-world applications. It would be very instructive to see these methods applied in an engineering project from initial design, optimization, data acquisition, calibration, field testing, validation, and up to production. Unfortunately, this was not possible within the scope of this thesis due to time and financial constraints.

Finally, research should be devoted to the application of the two developed methodologies to

the fields of Machine Learning and Scientific Machine Learning. Indeed, the various methods developed this last decades in ML and SciML produce empirical or semi-empirical models that also need to be calibrated and validated. Again, the idea is to tailor the calibration and the validation tasks toward the predictive objective of the model. It is our hope that this research will lead to more widespread acceptance of VVUQ methodologies to guide the development and use of these type of models in science and engineering.

REFERENCES

[1] K. J. Dowding, M. Pilch, and R. G. Hills, "Formulation of the thermal problem," *Validation Challenge Workshop*, vol. 197, no. 29, pp. 2385–2389, May 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S004578250700504X

[2] R. Ghanem, D. Higdon, and H. Owhadi, Eds., *Handbook of uncertainty quantification*, 1st ed. Springer Cham, 2017. [Online]. Available: https://doi.org/10.1007/978-3-319-12385-1

[3] T. Sullivan, *Introduction to Uncertainty Quantification*, ser. Texts in Applied Mathematics. Springer International Publishing, 2015, vol. 63.

[4] M. Ainsworth and J. T. Oden, *A posteriori error estimation in finite element analysis*, ser. Pure and applied mathematics. New York: Wiley, 2000.

[5] D. Estep and D. French, "Global Error Control For The Continuous Galerkin Finite Element Method For Ordinary Differential Equations," *RAIRO Modél. Math. Anal. Numér.*, vol. 28, Nov. 1994.

[6] S. Prudhomme and J. Oden, "On goal-oriented error estimation for elliptic problems: application to the control of pointwise errors," *Computer Methods in Applied Mechanics and Engineering*, vol. 176, no. 1-4, pp. 313–331, Jul. 1999. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0045782598003430

[7] J. Oden and S. Prudhomme, "Goal-oriented error estimation and adaptivity for the finite element method," *Computers & Mathematics with Applications*, vol. 41, no. 5-6, pp. 735–756, Mar. 2001. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0898122100003175

[8] K. Kergrene, S. Prudhomme, L. Chamoin, and M. Laforest, "A new goal-oriented formulation of the finite element method," *Advances in Computational Mechanics and Scientific Computationthe Cutting Edge*, vol. 327, pp. 256–276, Dec. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045782517306485

[9] D. S. Sivia and J. Skilling, *Data analysis: a Bayesian tutorial*, 2nd ed., ser. Oxford science publications. Oxford University Press, 2006.

[10] J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*, ser. Applied mathematical sciences. New York: Springer, 2005, no. v. 160.

[11] S. Sarkka, *Bayesian Filtering and Smoothing*. Cambridge: Cambridge University Press, 2013. [Online]. Available: http://ebooks.cambridge.org/ref/id/CBO9781139344203

[12] M. J. Schervish, *Theory of Statistics*, ser. Springer Series in Statistics. New York, NY: Springer New York, 1995. [Online]. Available: http://link.springer.com/10.1007/978-1-4612-4250-5

[13] A. DasGupta, *Asymptotic Theory of Statistics and Probability*, ser. Springer Texts in Statistics. Springer New York, 2008.

[14] H. White, "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, vol. 50, no. 1, pp. 1–25, 1982, publisher: [Wiley, Econometric Society]. [Online]. Available: http://www.jstor.org/stable/1912526

[15] B.J.K. Kleijn and A.W. van der Vaart, "The Bernstein-Von-Mises theorem under misspecification," *Electronic Journal of Statistics*, vol. 6, no. none, pp. 354–381, Jan. 2012. [Online]. Available: https://doi.org/10.1214/12-EJS675

[16] A. M. Stuart, "Inverse problems: A Bayesian perspective," *Acta Numerica*, vol. 19, pp. 451–559, May 2010. [Online]. Available: https://www.cambridge.org/core/product/identifier/S0962492910000061/type/journal_article

[17] J. Latz, "On the Well-posedness of Bayesian Inverse Problems," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 8, no. 1, pp. 451–482, Jan. 2020, publisher: Society for Industrial and Applied Mathematics. [Online]. Available: https://doi.org/10.1137/19M1247176

[18] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, 2001.

[19] J. Brynjarsdóttir and A. O'Hagan, "Learning about physical parameters: the importance of model discrepancy," *Inverse Problems*, vol. 30, no. 11, p. 114007, Oct. 2014, publisher: IOP Publishing. [Online]. Available: https://dx.doi.org/10.1088/0266-5611/30/11/114007

[20] C. Jiang, Z. Hu, Y. Liu, Z. P. Mourelatos, D. Gorsich, and P. Jayakumar, "A sequential calibration and validation framework for model uncertainty quantification and reduction," *Computer Methods in Applied Mechanics and Engineering*, vol. 368, p. 113172, Aug. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045782520303571

[21] N. Leoni, O. Le Maître, M.-G. Rodio, and P. M. Congedo, "Bayesian calibration with adaptive model discrepancy," *International Journal for Uncertainty Quantification*, vol. 14, no. 1, pp. 19–41, 2024.

[22] J. Hu, Q. Zhou, A. McKeand, T. Xie, and S.-K. Choi, "A model validation framework based on parameter calibration under aleatory and epistemic uncertainty," *Structural and Multidisciplinary Optimization*, vol. 63, no. 2, pp. 645–660, Feb. 2021. [Online]. Available: https://doi.org/10.1007/s00158-020-02715-z

[23] V. R. Joseph and S. N. Melkote, "Statistical Adjustments to Engineering Models," *Journal of Quality Technology*, vol. 41, no. 4, pp. 362–375, Oct. 2009, publisher: Taylor & Francis. [Online]. Available: https://doi.org/10.1080/00224065.2009.11917791

[24] Y. He and D. Xiu, "Numerical strategy for model correction using physical constraints," *Journal of Computational Physics*, vol. 313, pp. 617–634, May 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0021999116001261

[25] K. Sargsyan, H. N. Najm, and R. Ghanem, "On the Statistical Calibration of Physical Models," *International Journal of Chemical Kinetics*, vol. 47, no. 4, pp. 246–276, Apr. 2015, publisher: John Wiley & Sons, Ltd. [Online]. Available: https://doi.org/10.1002/kin.20906

[26] T. A. Oliver, G. Terejanu, C. S. Simmons, and R. D. Moser, "Validating predictions of unobserved quantities," *Computer Methods in Applied Mechanics and Engineering*, vol. 283, pp. 1310–1335, 2015.

[27] W. L. Oberkampf, T. G. Trucano, and C. Hirsch, "Verification, validation, and predictive capability in computational engineering and physics," *Applied Mechanics Reviews*, vol. 57, no. 5, pp. 345–384, Dec. 2004. [Online]. Available: https://doi.org/10.1115/1.1767847

[28] W. L. Oberkampf and T. G. Trucano, "Verification and validation benchmarks," *Benchmarking of CFD Codes for Application to Nuclear Reactor Safety*, vol. 238, no. 3, pp. 716–743, 2008.

[29] C. J. Roy and W. L. Oberkampf, "A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing," *Computer Methods in Applied Mechanics and Engineering*, vol. 200, no. 25, pp. 2131–2144, 2011.

[30] S. Ferson, W. L. Oberkampf, and L. Ginzburg, "Model validation and predictive capability for the thermal challenge problem," *Validation Challenge Workshop*, vol. 197, no. 29, pp. 2408–2430, 2008.

[31] R. Rebba and S. Mahadevan, "Computational methods for model reliability assessment," *Reliability Engineering & System Safety*, vol. 93, no. 8, pp. 1197–1207, 2008.

[32] J. Mullins, Y. Ling, S. Mahadevan, L. Sun, and A. Strachan, "Separation of aleatory and epistemic uncertainty in probabilistic model validation," *Reliability Engineering & System Safety*, vol. 147, pp. 49–59, 2016.

[33] Y. Ling and S. Mahadevan, "Quantitative model validation techniques: New insights," *Reliability Engineering & System Safety*, vol. 111, pp. 217–231, Mar. 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0951832012002414

[34] F. Llorente, L. Martino, D. Delgado, and J. López-Santiago, "Marginal Likelihood Computation for Model Selection and Hypothesis Testing: An Extensive Review," *SIAM Review*, vol. 65, no. 1, pp. 3–58, Feb. 2023, publisher: Society for Industrial and Applied Mathematics. [Online]. Available: https://doi.org/10.1137/20M1310849

[35] S. H. Cheung, T. A. Oliver, E. E. Prudencio, S. Prudhomme, and R. D. Moser, "Bayesian uncertainty analysis with applications to turbulence modeling," *Reliability Engineering & System Safety*, vol. 96, no. 9, pp. 1137–1149, Sep. 2011. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0951832011000664

[36] K. Farrell, J. T. Oden, and D. Faghihi, "A Bayesian framework for adaptive selection, calibration, and validation of coarse-grained models of atomistic systems," *Journal of Computational Physics*, vol. 295, pp. 189–208, 2015.

[37] K. Farrell-Maupin and J. T. Oden, "Adaptive selection and validation of models of complex systems in the presence of uncertainty," *Research in the Mathematical Sciences*, vol. 4, Aug. 2017, place: United States Publisher: SpringerOpen. [Online]. Available: https://www.osti.gov/pages/biblio/1356828

[38] J. Tan, B. Liang, P. K. Singh, K. A. Farrell-Maupin, and D. Faghihi, "Toward selecting optimal predictive multiscale models," *Computer Methods in Applied Mechanics and Engineering*, vol. 402, 2022.

[39] J. T. Oden and S. Prudhomme, "Control of modeling error in calibration and validation processes for predictive stochastic models," *International Journal for*

Numerical Methods in Engineering, vol. 87, no. 1-5, pp. 262–272, Jul. 2011, publisher: John Wiley & Sons, Ltd. [Online]. Available: https://doi.org/10.1002/nme.3038

[40] R. E. Morrison, C. M. Bryant, G. Terejanu, S. Prudhomme, and K. Miki, "Data partition methodology for validation of predictive models," *ICNC-FSKD 2012*, vol. 66, no. 10, pp. 2114–2125, Dec. 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0898122113005476

[41] A. Paquette-Rufiange, S. Prudhomme, and M. Laforest, "Optimal design of validation experiments for the prediction of quantities of interest," *Computer Methods in Applied Mechanics and Engineering*, vol. 414, p. 116182, Sep. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045782523003067

[42] R. G. Hills, M. Pilch, K. J. Dowding, J. Red-Horse, T. L. Paez, I. Babuka, and R. Tempone, "Validation Challenge Workshop," *Validation Challenge Workshop*, vol. 197, no. 29, pp. 2375–2380, May 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045782507005063

[43] G. J. Klir, *Uncertainty and information: foundations of generalized information theory.* Wiley-Interscience, 2006.

[44] B. Øksendal, *Stochastic Differential Equations*, ser. Universitext. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003. [Online]. Available: http://link.springer.com/10.1007/978-3-642-14394-6

[45] I. Babuka, F. Nobile, and R. Tempone, "A Stochastic Collocation Method for Elliptic Partial Differential Equations with Random Input Data," *SIAM Review*, vol. 52, no. 2, pp. 317–355, Jan. 2010. [Online]. Available: http://epubs.siam.org/doi/10.1137/100786356

[46] R. G. Ghanem and P. D. Spanos, *Stochastic finite elements: a spectral approach.* New York, NY: Springer, 1991, oCLC: 22733463.

[47] A. B. Owen, *Monte Carlo theory, methods and examples.* https://artowen.su.domains/mc/, 2013.

[48] G. B. Folland, *Real analysis: modern techniques and their applications*, 2nd ed., ser. Pure and applied mathematics. New York: Wiley, 1999.

[49] A. B. Owen, "Monte Carlo Variance of Scrambled Net Quadrature," *SIAM Journal on Numerical Analysis*, vol. 34, no. 5, pp. 1884–1910, Oct. 1997. [Online]. Available: http://epubs.siam.org/doi/10.1137/S0036142994277468

[50] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global Sensitivity Analysis. The Primer*. John Wiley & Sons, Ltd, 2007.

[51] A. Saltelli, K. Aleksankina, W. Becker, P. Fennell, F. Ferretti, N. Holst, S. Li, and Q. Wu, "Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices," *Environmental Modelling & Software*, vol. 114, pp. 29–39, Apr. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1364815218302822

[52] S. Razavi, A. Jakeman, A. Saltelli, C. Prieur, B. Iooss, E. Borgonovo, E. Plischke, S. Lo Piano, T. Iwanaga, W. Becker, S. Tarantola, J. H. Guillaume, J. Jakeman, H. Gupta, N. Melillo, G. Rabitti, V. Chabridon, Q. Duan, X. Sun, S. Smith, R. Sheikholeslami, N. Hosseini, M. Asadzadeh, A. Puy, S. Kucherenko, and H. R. Maier, "The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support," *Environmental Modelling & Software*, vol. 137, p. 104954, 2021.

[53] E. Borgonovo and E. Plischke, "Sensitivity analysis: A review of recent advances," *European Journal of Operational Research*, vol. 248, no. 3, pp. 869–887, Feb. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0377221715005469

[54] G. E. B. Archer, A. Saltelli, and I. M. Sobol, "Sensitivity measures,anova-like Techniques and the use of bootstrap," *Journal of Statistical Computation and Simulation*, vol. 58, no. 2, pp. 99–120, May 1997. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00949659708811825

[55] H. Rabitz and O. F. Aliş, "General foundations of high-dimensional model representations," *Journal of Mathematical Chemistry*, vol. 25, no. 2, pp. 197–233, Jun. 1999. [Online]. Available: https://doi.org/10.1023/A:1019188517934

[56] A. W. v. d. Vaart, *Asymptotic statistics*, ser. Cambridge series in statistical and probabilistic mathematics. Cambridge, UK ; New York, NY, USA: Cambridge University Press, 1998.

[57] I. Sobol, "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates," *Mathematics and Computers in Simulation*, vol. 55, no. 1, pp. 271–280, 2001.

[58] A. Puy, S. Lo Piano, and A. Saltelli, "Is VARS more intuitive and efficient than Sobol indices?" *Environmental Modelling & Software*, vol. 137, p. 104960, Mar. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364815221000037

[59] G. Chastaing, F. Gamboa, and C. Prieur, "Generalized Hoeffding-Sobol decomposition for dependent variables - application to sensitivity analysis," *Electronic Journal of Statistics*, vol. 6, no. none, Jan. 2012. [Online]. Available: https://projecteuclid.org/journals/electronic-journal-of-statistics/volume-6/issue-none/Generalized-Hoeffding-Sobol-decomposition-for-dependent-variables---application/10.1214/12-EJS749.full

[60] ——, "Generalized Sobol sensitivity indices for dependent variables: numerical methods," *Journal of Statistical Computation and Simulation*, vol. 85, no. 7, pp. 1306–1333, May 2015. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00949655.2014.960415

[61] A. B. Owen, "Sobol' Indices and Shapley Value," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 2, no. 1, pp. 245–251, Jan. 2014. [Online]. Available: http://epubs.siam.org/doi/10.1137/130936233

[62] M. Stein, "Large Sample Properties of Simulations Using Latin Hypercube Sampling," *Technometrics*, vol. 29, no. 2, pp. 143–151, May 1987. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00401706.1987.10488205

[63] I. Sobol and S. Kucherenko, "Derivative based global sensitivity measures and their link with global sensitivity indices," *Mathematics and Computers in Simulation*, vol. 79, no. 10, pp. 3009–3017, Jun. 2009. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0378475409000354

[64] M. Lamboni, B. Iooss, A.-L. Popelin, and F. Gamboa, "Derivative-based global sensitivity measures: General links with Sobol' indices and numerical tests," *Mathematics and Computers in Simulation*, vol. 87, pp. 45–54, Jan. 2013. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0378475413000141

[65] P. G. Constantine, E. Dow, and Q. Wang, "Active subspace methods in theory and practice: applications to kriging surfaces," *SIAM Journal on Scientific Computing*, vol. 36, no. 4, pp. A1500–A1524, 2014, arXiv: 1304.2070.

[66] P. G. Constantine, *Active subspaces: emerging ideas for dimension reduction in parameter studies*, ser. SIAM spotlights. Society for Industrial and Applied Mathematics, 2015, no. 2.

[67] P. G. Constantine and P. Diaz, "Global sensitivity metrics from active subspaces," *Reliability Engineering & System Safety*, vol. 162, pp. 1–13, Jun. 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0951832016303052

[68] P. G. Constantine, Z. del Rosario, and G. Iaccarino, "Data-driven dimensional analysis: algorithms for unique and relevant dimensionless groups," *arXiv:1708.04303 [math, stat]*, Aug. 2017, arXiv: 1708.04303. [Online]. Available: http://arxiv.org/abs/1708.04303

[69] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*, 3rd ed., ser. Adaptive computation and machine learning. Cambridge, Mass.: MIT Press, 2008.

[70] C. J. Paciorek and M. J. Schervish, "Spatial modelling using a new class of nonstationary covariance functions," *Environmetrics*, vol. 17, no. 5, pp. 483–506, Aug. 2006, publisher: John Wiley & Sons, Ltd. [Online]. Available: https://doi.org/10.1002/env.785

[71] M. D. Risser and C. A. Calder, "Regression-based covariance functions for nonstationary spatial modeling," *Environmetrics*, vol. 26, no. 4, pp. 284–297, Jun. 2015, publisher: John Wiley & Sons, Ltd. [Online]. Available: https://doi.org/10.1002/env.2336

[72] R. H. Byrd, J. Nocedal, and R. A. Waltz, "Knitro: An Integrated Package for Nonlinear Optimization," in *Large-Scale Nonlinear Optimization*, G. Di Pillo and M. Roma, Eds. Boston, MA: Springer US, 2006, pp. 35–59. [Online]. Available: https://doi.org/10.1007/0-387-30065-1_4

[73] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical Programming*, vol. 106, no. 1, pp. 25–57, Mar. 2006. [Online]. Available: http://link.springer.com/10.1007/s10107-004-0559-y

[74] C. Audet, S. L. Digabel, V. R. Montplaisir, and C. Tribes, "Algorithm 1027: NOMAD Version 4: Nonlinear Optimization with the MADS Algorithm," *ACM Transactions on Mathematical Software*, vol. 48, no. 3, p. 22, 2022.

[75] C. Audet and J. E. Dennis, "Mesh Adaptive Direct Search Algorithms for Constrained Optimization," *SIAM Journal on Optimization*, vol. 17, no. 1, pp. 188–217, Jan. 2006, publisher: Society for Industrial and Applied Mathematics. [Online]. Available: https://doi.org/10.1137/040603371

[76] J. Nocedal and S. J. Wright, *Numerical optimization*, 2nd ed., ser. Springer series in operations research. New York: Springer, 2006, oCLC: ocm68629100.

[77] C. Audet and J. E. Dennis, "A Progressive Barrier for Derivative-Free Nonlinear Programming," *SIAM Journal on Optimization*, vol. 20, no. 1, pp. 445–472, Jan. 2009, publisher: Society for Industrial and Applied Mathematics. [Online]. Available: https://doi.org/10.1137/070692662

[78] M. Peyrega, "Optimisation sans dérivées sous contraintes," Ph.D. dissertation, École Polytechnique de Montréal, 2016, phd. [Online]. Available: https://publications.polymtl.ca/2216/

[79] C. Audet, V. Béchard, and S. L. Digabel, "Nonsmooth optimization through Mesh Adaptive Direct Search and Variable Neighborhood Search," *Journal of Global Optimization*, vol. 41, no. 2, pp. 299–318, Jun. 2008. [Online]. Available: https://doi.org/10.1007/s10898-007-9234-1

[80] E. L. Lehmann, *Elements of Large-Sample Theory*. Springer, Dec. 1998.

[81] D. Cole, *Parameter redundancy and identifiability*. CRC Press, 2020.

[82] G. E. P. Box, A. Luceño, and M. d. C. Paniagua-Quiñones, *Statistical control by monitoring and adjustment*, second edition ed. Hoboken, N.J.: John Wiley & Sons, 2009. [Online]. Available: http://site.ebrary.com/id/10501391

[83] E. L. Lehmann and J. P. Romano, *Testing statistical hypotheses*, 3rd ed., ser. Springer texts in statistics. New York: Springer, 2005.

[84] A. White, S. Mahadevan, J. Schmucker, and A. Karl, "Multi-Metric Validation Under Uncertainty for Multivariate Model Outputs and Limited Measurements," *Journal of Verification, Validation and Uncertainty Quantification*, vol. 7, no. 041004, Jan. 2023. [Online]. Available: https://doi.org/10.1115/1.4056548

[85] S. Oladyshkin and W. Nowak, "The Connection between Bayesian Inference and Information Theory for Model Selection, Information Gain and Experimental Design," *Entropy*, vol. 21, no. 11, 2019.

[86] S. Riedmaier, B. Danquah, B. Schick, and F. Diermeyer, "Unified Framework and Survey for Model Verification, Validation and Uncertainty Quantification," *Archives of Computational Methods in Engineering*, vol. 28, no. 4, pp. 2655–2688, 2021.

[87] G. Lee, W. Kim, H. Oh, B. D. Youn, and N. H. Kim, "Review of statistical model calibration and validationfrom the perspective of uncertainty structures," *Structural and Multidisciplinary Optimization*, vol. 60, no. 4, pp. 1619–1644, 2019.

[88] J. R. Hamilton and R. G. Hills, "Relation of Validation Experiments to Applications," *Numerical Heat Transfer, Part B: Fundamentals*, vol. 57, no. 5, pp. 307–332, 2010.

[89] R. G. Hills, "Roll-up of validation results to a target application." Sandia National Laboratories, United States, Tech. Rep. SAND2013-7424, 2013.

[90] R. G. Hills and I. H. Leslie, "Statistical Validation of Engineering and Scientific Models: Validation Experiments to Application," Sandia National Laboratories, United States, Tech. Rep. SAND2003-0706, 2003.

[91] C. Li and S. Mahadevan, "Role of calibration, validation, and relevance in multi-level uncertainty integration," *Reliability Engineering & System Safety*, vol. 148, pp. 32–43, 2016.

[92] D. Ao, Z. Hu, and S. Mahadevan, "Design of validation experiments for life prediction models," *Reliability Engineering & System Safety*, vol. 165, pp. 22–33, 2017.

[93] I. Sunseri, J. Hart, B. van Bloemen Waanders, and A. Alexanderian, "Hyper-differential sensitivity analysis for inverse problems constrained by partial differential equations," *Inverse Problems*, vol. 36, no. 12, p. 125001, 2020.

[94] F. M. White, *Fluid mechanics*, 7th ed., ser. Mcgraw-Hill series in mechanical engineering. McGraw-Hill, 2009.

[95] S. Da Veiga, F. Gamboa, B. Iooss, and C. Prieur, *Basics and Trends in Sensitivity Analysis*, ser. Computational Science & Engineering. Society for Industrial and Applied Mathematics, 2021.

[96] P. D. Arendt, D. W. Apley, and W. Chen, "Quantification of Model Uncertainty: Calibration, Model Discrepancy, and Identifiability," *Journal of Mechanical Design*, vol. 134, no. 10, 2012.

[97] J. G. Mullins, B. B. Schroeder, R. G. Hills, and L. G. Crespo, "A Survey of Methods for Integration of Uncertainty and Model Form Error in Prediction." in *Report Number: SAND2016-4585C*, 2016, sandia National Laboratories, United States.

[98] J. Goodman and J. Weare, "Ensemble samplers with affine invariance," *Communications in Applied Mathematics and Computational Science*, vol. 5, no. 1, pp. 65–80, 2010.

[99] A. C. Atkinson, "Optimal Design," in *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd, 2015, pp. 1–17.

[100] A. C. Atkinson, A. N. Donev, and R. D. Tobias, *Optimum experimental designs, with SAS*, ser. Oxford statistical science series. Oxford University Press, 2007, no. 34.

[101] J. Beck, B. M. Dia, L. F. Espath, Q. Long, and R. Tempone, "Fast Bayesian experimental design: Laplace-based importance sampling for the expected information gain," *Computer Methods in Applied Mechanics and Engineering*, vol. 334, pp. 523–553, 2018.

[102] Q. Long, M. Scavino, R. Tempone, and S. Wang, "Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations," *Computer Methods in Applied Mechanics and Engineering*, vol. 259, pp. 24–39, 2013.

[103] E. G. Ryan, C. C. Drovandi, J. M. McGree, and A. N. Pettitt, "A Review of Modern Computational Algorithms for Bayesian Optimal Design," *International Statistical Review*, vol. 84, no. 1, pp. 128–154, 2016.

[104] H. Yang, M. Fan, A. Liu, and L. Dong, "General formulas for drag coefficient and settling velocity of sphere based on theoretical law," *International Journal of Mining Science and Technology*, vol. 25, no. 2, pp. 219–223, 2015.

[105] S. Razavi and H. V. Gupta, "A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory," *Water Resources Research*, vol. 52, no. 1, pp. 423–439, 2016.

[106] M. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. Rognes, and G. Wells, "The FEniCS project version 1.5," *Archive of Numerical Software*, vol. 3, 2015.

[107] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, Eds., *Optimization with PDE constraints*, ser. Mathematical modelling: theory and applications. Springer, 2009, no. 23.

[108] M. S. Alnæs, A. Logg, K. B. Ølgaard, M. E. Rognes, and G. N. Wells, "Unified Form Language: A Domain-Specific Language for Weak Formulations of Partial Differential Equations," *ACM Trans. Math. Softw.*, vol. 40, no. 2, 2014.

[109] A. Montoison, P. Pascal, and L. Salomon, "NOMAD.jl: A Julia interface for the constrained blackbox solver NOMAD," 2020. [Online]. Available: https://github.com/bbopt/NOMAD.jl