Characterization of the impact of material variability on the parameters of composite behavior laws

Clément Laboulfie

NNT : 2023EMSEM038

THÈSE DE DOCTORAT

de Mines Saint-Etienne - Une école de l'IMT

Ecole Doctorale N°488
(Sciences, Ingénierie, Santé)

Spécialité de doctorat : Sciences pour l'ingénieur DS8

Soutenue publiquement le 16/11/2023 par :

Clément LABOULFIE

Caractérisation de l'impact de la variabilité matériau sur les paramètres de lois de comportement composites

Devant un jury composé de :

CHAMOIN Ludovic	Professeur, ENS Paris-Saclay	Rapporteur
GOGU Christian	Professeur, ISAE-Supaéro	Rapporteur
CHEVREUIL PLESSIS Mathilde	Maître de Conférences, Nantes Université	Examinatrice
GUILLAUMAT Laurent	Professeur, ENSAM Angers	Examinateur
LE RICHE Rodolphe	Directeur de Recherche, CNRS LIMOS	Directeur de thèse
IRISARRI François-Xavier	Directeur de Recherche ONERA/DMAS	Co-encadrant
BREVAULT Loïc	Ingénieur de Recherche ONERA/DTIS	Co-encadrant

Affidavit

Je soussigné, Clément LABOULFIE, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de Rodolphe LE RICHE, dans le respect des principes d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect de la charte nationale de déontologie des métiers de la recherche.

Ce travail n'a pas été précédemment soumis dans sa globalité en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Châtillon., le 11 septembre 2023

REMERCIEMENTS

Au-delà de sa seule dimension scientifique, un doctorat est avant tout une aventure humaine. Dès lors, nulle surprise qu'un doctorat ne soit rempli de moments de fatigue ou de découragement, qui, il faut le souligner, laissent quelques fois la place, au détour d'une ligne de code ou d'une équation, à une exaltation digne des épopées d'antan. Par ailleurs, soulignons que le travail de recherche est tout sauf solitaire et que si parfois les affres de la gestion d'équipe peuvent décourager l'ermite, elles figurent pour le meilleur et pour le pire au cœur du travail de recherche. En ce qui me concerne, j'ai été, pendant cette odyssée, aidé par des argonautes d'exception. Je voudrais ici commencer par adresser mes plus vifs remerciements à mon encadrement (Rodolphe Le Riche, Mathieu Balesdent, Loïc Brévault, François-Xavier Irrisari, Jean-François Maire et Sébastien Da Veiga), au vu de leur dévouement, de leur implication, mais aussi de l'intérêt constant qu'ils ont porté à mes travaux de recherche. En outre, j'ai été particulièrement sensible à la réactivité, aux échanges, conseils et remarques que MM. Brévault & Balesdent m'ont prodigués et qui ont contribué à former le chercheur que je suis devenu et, je n'en doute pas, à faire progresser celui que je deviendrai. Je n'oublie pas non plus Christian Fagiano, directeur du projet PRINCAST dans lequel ma thèse s'inscrivait ; il m'a, au cours de ces années de travail, dispensé des soutiens amicaux et efficaces, pimentés de nos joutes verbales enflammées sur les performances sportives des Squadre Azzure, rugby et foot compris. Au sein de l'équipe qui m'a permis d'atteindre le titre de docteur, je voudrais aussi remercier l'ensemble des doctorants, Ludovic, Claire, Mathieu, Loïc avec qui je partageais le bureau, ainsi que Stacy, Arthur, Mathis, Enrico, Antoine, Gaspard, Pierre, Florent et Salim. Impossible d'oublier ces parties de UNO acharnées, rythmées par des interceptions endiablées et qui ont constitué des moments de convivialité que l'on souhaiterait retrouver partout ailleurs. Je salue aussi plus généralement l'équipe MC^2, au sein de laquelle j'ai eu le plaisir de réaliser mes travaux de thèse. Je pense aussi au soutien et à la confiance que mes parents et ma sœur ont pu me donner au cours de ces années, en particulier pendant ces moments difficiles où on ne semble plus toucher terre. Comment ne pas les remercier du fond du cœur pour leur engagement total et infaillible. Je voudrais terminer ces remerciements en adressant une dédicace spéciale à François-Henri Leroy, avec qui j'ai pu avoir des discussions franches, passionnées et passionnantes qui font, de mon point de vue, tout le sel de la recherche. Je souhaite à tous ceux qui s'aventurent sur le chemin de la recherche de pouvoir éprouver le même plaisir, le même frisson d'excitation, que j'ai ressenti lorsque j'ai appris que j'étais autorisé à poursuivre le stage dans le cadre d'une thèse. Je remercie une dernière fois toutes les personnes qui, de près ou de loin, m'ont permis de réaliser avec succès ce projet de thèse.

CONTENTS

GENERAL INTRODUCTION

In response to climate change, environmental regulation agencies such the US Environmental Protection Agency are imposing limits on greenhouse gas emissions [EPA, 2021] for the aeronautic industry. These constraints add to the safety requirements that must still be met in order to obtain an operating license. These requirements have to be taken into account in an increasingly competitive aviation industry. To comply with these new environmental rules, one of the available solution consists in reducing the weight of the aircrafts while maintaining the safety constraints: less weight means less fuel consumption, which in turn reduces greenhouse gas emissions. Practical solutions to decrease the weight of aeronautical structures include replacing metallic structures by composites materials.

Composites materials can be defined as "the combination of two materials whose joint mechanical performance is better than when considered individually" [Berthelot, 2012]. Natural examples of composite materials include wood. There are many types of composite materials ([Rana and Fangueiro, 2016]) that differ either in the constituents (e.g. Carbon or Glass fibers and polymer, metal or ceramic matrix) or in the shape (short, long or continuous fibers) and the arrangement (woven or unidirectionnal composites) of the reinforcements. Compared to their metallic counterparts, one of their main advantages is their lower densities which is the reason why they are used by the aerospace industry. In average, a decrease of 30% of the weight of the overall structure can be expected [Rana and Fangueiro, 2016]. This is illustrated with the Boeing 787 Dreamliner, composed of approximately 50% of composite materials and whose weight has been decreased by 20% compared to conventional aluminum designs, according to the manufacturer [Hale, 2006]. Aside from these considerations, composite materials have several other advantages. These include high resistance to fatigue, [Hale, 2006, Rana and Fangueiro, 2016]. Furthermore, thanks to a reduction in the number of parts, they may allow to speed up the manufacturing process, resulting in a higher production rate (from 11 months with metallic designs to only 7 months with composite materials to assemble an aircraft [Rana and Fangueiro, 2016]). Finally, unlike most metals, they are not subject to corrosion problems, which reduces maintenance costs (about 35% fewer working hours for the tail of the Boeing 777 compared to the Boeing 767 aluminum tail, according to the manufacturer [Hale, 2006]).

By definition, composite materials are made from at least two different materials [Berthelot, 2012]. Usually, one of them corresponds to the reinforcement (in our case, continuous fibers) and the other to the resin. Their properties can have similar values (it is for instance the case for some carbon/carbon composites) or exhibit significant differences as for the standard carbon/epoxy composites used in aeronautical structures. In fact, most of the time, the fibers demonstrate a greater resistance than the resin, in particular for the failure properties (see for instance the data sheet for T700GCM21 material [Hexcel, 2020]). Aside from their strength

properties, the constituents can also have different behaviors. For instance, the resin may have a viscous behavior in contrast to the fibers. The combination of the two constituents results in an heterogeneous material whose homogenized mechanical behavior presents a significant anisotropy. In other words, the material properties vary depending on the direction of observation, contrary to their isotropic counterparts.
In addition to their heterogeneities, composite materials are also multi-scale materials. Usually, the constituents scale is called the micro-scale. The scale of the composite structures is called the macro-scale. For laminated composites, an intermediate scale called the meso-scale referring to the base ply is introduced. These scales are depicted in Figure 1.

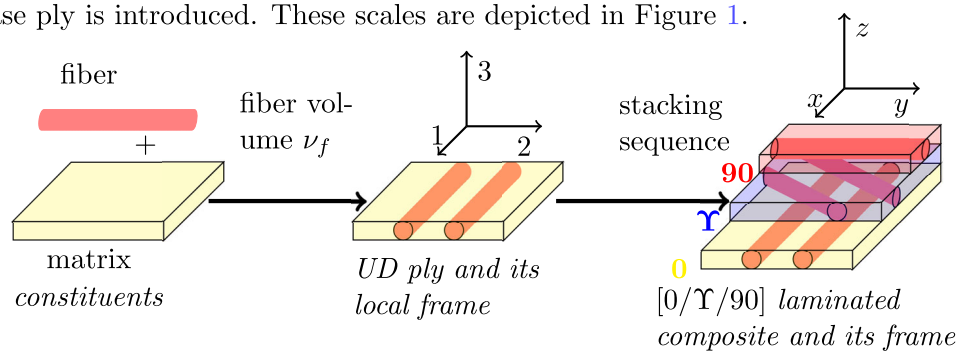


Figure 1: From left to right: micro, meso and macro-scales for laminated composite materials. Here, the base ply is unidirectionnal (UD) with all fibers aligned in the same direction. Υ stands for the orientation of the mid ply.

The properties of industrial interest are those of the base ply determined on experiments conducted at the macro-scale, in particular because experiments are easier to conduct at larger scales.
The heterogeneity and multi-scale aspects of composite materials lead to a major feature: the properties of such materials are subject to significant variability. The uncertainty that can have a significant impact on the meso-scale properties is the fiber volume fraction which is considered as the most significant sources of variability on the elastic properties of the base ply [Smarslok et al., 2012]. Other sources of uncertainties such as those that concern fiber misalignment [Gusev et al., 2000, Fougerouse, 2023] can also be taken into account to get a better modeling of the variability of the material properties. The uncertainties at the micro-scale combine with those related to the meso-scale. Among them can be mainly found the misalignment of the plies [Chen et al., 2017] or their thickness over or under-estimation. Such phenomena impact not only the elastic behavior but also the failure behavior (see [Rollet, 2007] for the energy release rate). Note also that, aside from the base materials, the manufacturing process itself brings some variability to the material properties (see [Rollet, 2007]).
To ensure a satisfying level of safety, the certification process should account for the main specifities of the composites behavior spotted above. For instance, the impact of the multi-scale behavior is characterized by performing mechanical tests on samples of different sizes from the base samples up to complete structures, as depicted in Figure 2.

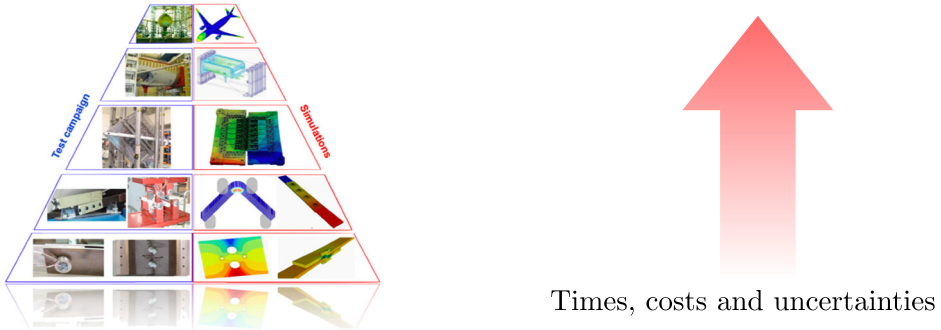


Figure 2: Different types of tests required for the certification of aeronautic structures (with the courtesy of F. Laurin).

To ensure that the designed structures are resistant to external and internal solicitations in most, if not all, cases and meet the safety requirements established by the regulatory authorities, the impact of the intrinsic variability of the material properties on the design allowables (*e.g.*, ultimate stress, buckling load) should be estimated. In order to identify the impact of this intrinsic variability on the material properties, international standards require that the certification tests should be repeated several times [Defense et al., 1999]. Of course, as shown in Figure 2, as the size of the tested specimen increases, the corresponding experiments become more difficult and costly to perform due to practical problems. Consequently, in the large scales that are of practical interest, the measured variability can be challenged because of the lack of test repetitions (on the complete structures, only one test is achieved). A solution is to apply deterministic safety factors to the design criteria to ensure the absence of any issue with, for instance, the use of A or B-values which characterize the 1 and 10% quantiles with 95% confidence [Defense et al., 1999]. Yet, such a solution generally implies to oversize the structure that will meet and even outperform safety requirements, at the expense of undesirable extra costs and weights [Beck and de Santana Gomes, 2012].

A second strategy illustrated in [Lekou and Philippidis, 2008, Rollet, 2007], based on uncertainty propagation and numerical simulations, allows to strike a better balance between safety, weights and costs constraints and applies in two main steps. It consists, first, in characterizing material variability using small-scales experiments easier to perform in large numbers. Then, using uncertainty quantification techniques such as Monte-Carlo based methods [Metropolis and Ulam, 1949], the impact of the variability of the material properties can be propagated across the scales to approximate the dispersion of the design criterion of interest. A cornerstone of this strategy is the use of numerical simulations to predict the material response.
To perform these numerical studies, it is necessary to run a simulator that mimics the material behavior which can be defined through a mathematical relation called a behavior law. To be more precise, a behavior law relates the stress $\boldsymbol{\sigma}$ (in MPa) to the dimensionless strain $\boldsymbol{\varepsilon}$ (or their derivative versus time) *via* a relation that usually reads as:

$$\boldsymbol{\sigma}(t, \boldsymbol{\theta}) = \underline{\underline{\mathbf{C}}}(t, \boldsymbol{\theta}) : \varepsilon_{\text{elas}}(t, \boldsymbol{\theta}) \ , \tag{1}$$

with $\underline{\underline{\mathbf{C}}}$ the stiffness tensor and $\varepsilon_{\text{elas}}$ the elastic part of the strain. Such behavior laws call for model parameters labeled $\boldsymbol{\theta}$ that will be discussed further below. The dimension of the quantities depends on the model of interest. In fact, most of the complexity of the behavior law resides in the expression of the stiffness tensor and the formulation of the elastic strain as a function of the observed strain $\boldsymbol{\varepsilon}$. For instance, damage variables d can be added to the stiffness tensor. If in the simplest case the elastic strain corresponds to the total strain, more complex expressions involving viscous strains (that depend on the model parameters and time) can be employed to represent the nonlinearitites of the material behavior (hence the dependence on variables $\boldsymbol{\theta}$ and t) [Berthelot, 2012]. In this work, a focus will be made on simplified versions of two behavior laws developed at ONERA: the ONERA Damage Model (ODM) [Maire et al., 1997, Marcin et al., 2011, Ben Ramdane, 2014] for woven composites and the ONERA Progressive Failure Model (OPFM) [Laurin, 2005] for laminates. Note that because of the anisotropy of the material, the most complex laws can have many model parameters (about 21 for the OPFM model).

The model parameters $\boldsymbol{\theta}$ introduced in Eq.(1) are of prime importance in defining the material behavior. In fact, they characterize the material behavior together with the mathematical modeling and the same strain-stress relation can provide different responses for significantly different model parameters $\boldsymbol{\theta}$, all other settings being equal. It should be highlighted that there are no default values for $\boldsymbol{\theta}$ which depend on the studied material. Consequently, they should be estimated or rather "calibrated" before performing any simulations. Model calibration is a long-standing issue in material science. Briefly, the problem to solve is formulated as follows:

given input data (for instance a strain profile $\boldsymbol{\varepsilon}_{\text{exp}}$), the corresponding output data (for instance the stress measurements associated to $\boldsymbol{\varepsilon}_{\text{exp}}$) observed on a material that is supposed to be described by a behavior law (let take for example the one in Eq. (1)), what is the value of $\boldsymbol{\theta}$ for which the simulated model response "best" fit the observations [Banks et al., 2014, Tarantola, 2005] ? Even the meaning of "best" needs to be specified through the formulation of a metric. Many methods have been developed to infer the parameters among which can be found the popular least square methods [Banks et al., 2014, Tarantola, 2005] or those based on likelihood maximization [Pintelon and Schoukens, 2012, Pollak and Palazotto, 2009, Wasserman, 2010]. However, a deterministic estimate of the model parameters is not fully satisfactory. Indeed, for instance, the observations are flawed with measurement errors [Kennedy and O'Hagan, 2001, Tarantola, 2005]. Furthermore, behavior laws are not perfect because they do not capture properly all the physics that would enable a perfect description of the material behavior, introducing model bias [Kennedy and O'Hagan, 2001]. Lots of other types of uncertainties can be also considered [Kennedy and O'Hagan, 2001] such as, for example, code uncertainty. Therefore, whatever uncertainties are considered relevant, to provide reliable predictions, it is necessary to consider their impact on the model parameters $\boldsymbol{\theta}$. Many techniques have been developed to quantify the uncertainties on the model parameters, that can be generally divided into two main branches: the frequentist methods [Gallagher and Doherty, 2007] and the Bayesian framework [Kennedy and O'Hagan, 2001]. Frequentist methods include the asymptotic theory [Antoniadis and Carmona, 1992, Banks et al., 2014, Gallagher and Doherty, 2007] and bootstrap-based techniques [Efron, 1979]. Both types of calibration methods aim at characterizing the uncertainties on the model parameters with probability distributions, that can be sampled to perform uncertainty propagation. Standard calibration methods have been developed to calibrate behavior laws considering a unique test. However, remember that in our context, the material properties are subjected to an intrinsic variability that can be observed on repetitions of experiments and cannot be identified on a single test. In fact, when considering repetitions of experiments, an implicit hypothesis of most standard calibration approaches is that the repeated specimens can be described by a single set of model parameters $\boldsymbol{\theta}$, suggesting that intrinsic variability is in fact negligible. This is not consistent with experimental observations on composites [Rollet, 2007] and leads to the following conclusion: such methods do not seem to be designed to characterize material variability, simply because their key assumptions state that it is negligible. It should be underlined that the variability of some properties can be directly measured from observations as for the failure stresses on strain-stress curves or the damage thresholds from acoustic data among others [Maljaars et al., 2022]. Yet, this only concerns a relatively small number of parameters and others require a calibration step to be estimated.

Following these observations, the objective of this thesis is to set up a calibration method that would account for measurement errors, model inadequacy (neglected for the sake of simplicity) and in particular material variability. This research was co-financed by ONERA and Safran-Tech, whose ambition consists in characterizing the impact of intrinsic variability of composites on the material properties to reduce design costs while maintaining satisfying levels of safety. To be more precise, given a behavior law on the one hand and repetitions of experiments on the other hand, the goal is to design a calibration process to calibrate the model in the presence of uncertainties and to determine the impact of material variability on the model parameters. Because of its practical uses by material engineers, the estimated variability should be modeled with a joint probability distribution. With such a tool, many different investigations based on uncertainty propagation can be achieved: sensitivity analyses, estimation of design quantities (*e.g.*, failure stresses, buckling loads, failure probabilities), optimization under uncertainties, *etc.* Moreover, a joint distribution can account for the dependencies between the model parameters. These dependencies attest that physical phenomena are interacting. They are of prime importance as neglecting them can lead to erroneous computations of design quantities [Smarslok

et al., 2012].
The framework of the population approaches [Demidenko, 2013, Lavielle, 2014] addresses many of the requirements listed above. It is in fact an extension of the standard identification methods that explicitly incorporates the variability of the mechanical properties in its key assumptions. In practice, the estimation of the variability can be achieved in either the Bayesian or the frequentist frameworks. Such methods are widely used in many fields, in particular pharmacokinetics [Lavielle, 2014] and civil engineering [Song et al., 2020]. However, most of the current applications of the population approaches suffer from several limitations that make them difficult to implement in our context. A first limitation is the numerical complexity of the behavior laws used in mechanical engineering. Indeed, contrary to lots of models used in pharmacokinetics, the simulators cannot be expressed in closed form. They can require the resolution of non-linear differential equations and their output can be multivariate, which means that the model response is not a scalar but a vector. In addition, because of the complexity of their formulation, a model run can take a long time to provide the simulated response. Finally, it should be noted that to ensure the reliability of the results, it is necessary to have as much individuals as possible. However, because of financial limitations, it is not always possible to have many specimens available. It is therefore necessary to study the application of such methods to mechanical problems in order to analyze their ability to estimate the intrinsic variability of mechanical properties from a limited repetition of specimens. A preliminary study on these matters, deepened in this manuscript with in particular a comparison of several numerical methods and an inquiry on the impact of the number of specimens on the convergence of the estimation of the variability of the material properties, has been presented at the UNCECOMP 2021 conference in Athens in a talk entitled:
"Laboulfie, C., Balesdent, M., Brevault, L., Da Veiga, S., Irisarri, F.X., Le Riche, R. and Maire, J.F., 2021, June. Calibration of material model using mixed-effects models. In 4th International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2021) [Laboulfie et al., 2021]."
Another complication comes from the large number of model parameters for the behavior laws in mechanical engineering. Remember that those models can have up to 20 model parameters such as for the OPFM model [Laurin, 2005]. Given that most classical random variable distributions are parameterized by at least two parameters (*e.g.*, mean and standard deviation), this would make 40 population parameters to estimate, which is challenging from a numerical point of view. In addition, it should be noted that these different parameters do not have the same impact on the material response depending on the type of experiments. For instance, viscosity does not impact much the longitudinal behavior, and thus the parameters related to this phenomena will be difficult to estimate from these experiments. In general, specific experiments are carried out to isolate the different phenomena of interest. For instance, the longitudinal elastic coefficient is calibrated from tension/compression experiments on uni-directional 0° laminates. Taking advantage of this separation allows to consider a sequential strategy that aims to replace the full calibration problem by sub-problems of smaller dimensions (thus easier to solve). Such a sequential calibration approach has already been developed for standard calibration problems [Oberkampf and Roy, 2010, Urbina et al., 2012, Youn et al., 2011]. However, its adaptation to mixed-effects has not been considered yet. The development of a sequential process compliant with the population approaches is thus of prime importance in order to calibrate complex behavior laws using mixed-effects approach. This point has been introduced in a conference paper at the CFM meeting in August 2022 in Nantes (see [Laboulfie et al., 2022]):
"Laboulfie, C., Balesdent, M., Brevault, L., Da Veiga, S., Irisarri, F.X., Le Riche, R. and Maire, J.F., 2022, August. Sequential calibration of material model using mixed-effects. In 25ème Congrès Français de la mécanique (CFM 2022) [Laboulfie et al., 2022]."
It has further been developed in a journal paper and published in Mechanics & Industry : "Laboulfie C., Balesdent M. , Brevault L., Maire J.-F., Da Veiga S., Le Riche R., Sequential

calibration of material constitutive model using mixed-effects calibration, Mechanics & Industry **24**, 32 (2023) (2023) [Laboulfie et al., 2023] ".

We will now present the different steps of this work. Chapter 1 begins with a literature review of the standard calibration methods. It is followed by an application of these approaches to repetitions of tensile tests for a woven composite material (CERASEP A400 [Naslain, 2004, Viricelle et al., 2001]) to highlight the presence of material variability and perform a first estimation with these methods. Several objectives are pursued. First, it allows to test the settings of the different calibration methods and compare their respective advantages and drawbacks. Second, it enables to become familiar with the type of model we aim to identify, to discover the experimental data and analyze potential associated difficulties. The final objective is to illustrate the impact of material variability and the difficulties to estimate it properly with standard calibration techniques.
In the second chapter 2, the population modeling is presented, with a focus on mixed-effects models, an estimation approach for the estimation of material variability from replications of experiments.
In the third chapter 3, the mixed-effects methodology is implemented to investigate its applicability and discuss its settings. An in-depth investigation of the method hypotheses is conducted on virtual experiments. A focus is made on the impact of the number of available test repetitions, which proves to be a key feature in population-based approaches. To be more precise, the interest of considering complex model parameter distributions (that can include statistical dependencies or non Gaussian marginals for instance) is inquired. Furthermore, the experimental data treated in the first chapter is processed with mixed-effects to test the ability of the method to characterize material variability. This phase is of major interest because, contrary to the synthetic data, experimental data suffers from many "drawbacks". For instance, as it is not directly generated from the model, the behavior law cannot catch all the physics. The estimated variabilities are propagated through the behavior law. Recommendations can be set about the interest of considering complex assumptions on the model parameter distribution.
Finally, Chapter 4 proposes a sequential strategy to handle the calibration of different types of experiments. Emphasis is put on the accuracy of the results and on the computational costs implied by the method compared to the standard mixed-effects implementation. The proposed approach is then applied to repetitions of experiments carried out on a composite material with carbon fibers and epoxy resin. The results are compared to those coming from the standard application of mixed-effects. This implementation permits to assess the interest of the method on real data. It also enables to analyze the limitations of the mechanical modeling. Given that there are no exact references to which the estimated variability can be compared, a relevant way to discriminate between the results is to propagate the estimated distributions to approximate the distribution of the failure stress of a perforated plate of composite material. This illustrates the whole process of uncertainty quantification, which starts with the estimation of the variability of the material properties and ends with the propagation of the joint distribution of the material properties to estimate a design criterion (the failure stress of a perforated plate in tension). Several mechanical assumptions are also discussed in this uncertainty propagation.

CHAPTER 1

ESTIMATING MATERIAL VARIABILITY WITH STANDARD UNCERTAINTY QUANTIFICATION METHODS

Contents

Chapter goals:

- Introduction of model calibration and the importance of uncertainty quantification
- Description of the notations adopted in this manuscript
- Setting up of the statistical framework and reporting the assumptions chosen in this manuscript
- List of the methods implemented for deterministic model calibration with one or several replications of experiments
- Reporting the methods to carry out uncertainty quantification in either the Bayesian or frequentist frameworks
- Implementation the standard calibration methods on 13 replications of tensile tests performed on Ceramic Matrix Composite material
- Highlight of material variability and try to characterize it using multi-objective methods

1.1 Introduction

The quantification of uncertainty is a major issue in material science, especially when applied to the design of complex composites structures. Until now, most of the available procedures rely on safety factors associated to deterministic margins of safety [Beck and de Santana Gomes, 2012], for instance on strength properties. Though these margins provide satisfactory levels of reliability, most of the time, they result in the over-sizing of the structure, which leads to undesirable costs [Beck and de Santana Gomes, 2012]. A complete uncertainty quantification enables to design structures compliant with both costs and safety constraints (*e.g.*, constraints about buckling, about the strength of the structure). This quantification can rely on experiments that should be conducted in sufficient number to ensure consistency of the results. Yet, this solution implies both large durations and financial costs. In addition, the impact of measurements errors also increases with the size of the tested structure. Consequently, a recent trend of research consists in substituting numerical simulations to some experiments. These simulations require the development of numerical mechanical models that can be representative of the experimental data. Though they may require significant numerical facilities, they can model several phenomena. Such mechanical models are parameterized by model parameters (Young's modulus, Poisson's ratio, *etc.*) that need to be estimated before being used. Given the presence of measurement errors and possible inaccuracies of the mechanical model, to get reliable predictions, it is necessary to take into account the impact of the uncertainties on the model parameters. In model calibration, usual uncertainties refer to experimental noise and model bias [Kennedy and O'Hagan, 2001]. Many techniques have been developed to quantify their impact [Kennedy and O'Hagan, 2001, Banks et al., 2014] on model parameters and are now widely applied [Rappel et al., 2019a]. This work aims to characterize the impact of uncertainties on parameters of mechanical models dedicated to composites material. Such materials are known to be subjected to significant intrinsic variability due to the production process for instance [Defense et al., 1999]. To estimate this intrinsic variability, international standards impose to perform replications, that is to say to repeat the test with different samples from the same material. However, in usual model calibration, intrinsic variability is mostly neglected and combined with other sources of uncertainty. It is not a problem as long as it remains negligible but may create issues if it is critical as in the present case.

This chapter first aims to introduce the standard uncertainty quantification methods used in model calibration in a literature review. In a second time, such methods are applied on a mechanical test-case to illustrate that they are not designed to characterize the observed variability over repetitions of experiments. Section 1.2 describes the statistical framework and the different assumptions that can be made along with their consequences on the calibration method. Second, the main technique for deterministic model calibration with one or several experiments is described in Section 1.3. Then, the common methods to perform uncertainty quantification, the frequentist or Bayesian approaches, are summarized in Sections 1.4 and 1.5. Considering the computational costs required by such methods, techniques to speed up the calibration process are briefly discussed in Section 1.6. Finally, after the literature review, the presented methods are applied in Section 1.7 to a mechanical test case to illustrate their behavior with respect to the estimation of material variability.

1.2 Formalization of behavior law calibration

1.2.1 Notations

Let us first describe the notations and the main hypotheses that will be used throughout this manuscript. The probability space will be noted $(\Gamma_x, \mathcal{T}_x, \mathbb{P}_x)$ with Γ_x the set of possible values representing the physical and mathematical constraints, $\mathcal{T}_x$ the sigma-algebra and $\mathbb{P}_x$ the probability measure for any quantity x. This allows to define the random variable X. All multi-dimensional values (vector, matrix, tensor, *etc.*) will be written in bold type. The behavior law we seek to calibrate is labeled $\mathcal{F}(\cdot, \boldsymbol{\theta})$ and parameterized by the model parameters $\boldsymbol{\theta} \coloneqq [\theta_1, \dots, \theta_d] = [\theta_l]_{l \in [\![1,d]\!]} \in \mathbb{R}^d$ (d denotes the number of parameters to estimate and $l \in [\![1, d]\!]$ is the corresponding index). More generally, the vectors of parameters to be calibrated will be denoted between brackets, $[\dots]$. Its corresponding random vector is $\boldsymbol{\Theta}$. Concatenation of outcomes of random variables will be noted between braces. For instance, the collection of 4 realizations of $\boldsymbol{\Theta}$ will be noted $\{\boldsymbol{\theta}_i\}_{i \in [\![1,4]\!]}$. The objective is to mimic, with the behavior law, the experimental output data $\mathbf{y} \coloneqq (y_j)_{j \in [\![1,N]\!]} \in \mathbb{R}^N$ with N the number of observations and $j \in [\![1, N]\!]$ the corresponding index. The associated random variable is $\mathbf{Y}$. Unless specified otherwise, the observed output $\mathbf{y}$ is a vector of scalar quantities. It is associated with scalar deterministic input data $\mathbf{t} \coloneqq (t_j)_{j \in [\![1,N]\!]}$. More generally, the coordinates of temporal data will be denoted between parentheses, $(\dots)$. The set of real matrices of p rows by q column will be denoted $\mathcal{M}_{pq}(\mathbb{R})$. If $p = q$, it is simply written $\mathcal{M}_p(\mathbb{R})$. Sequences of elements x_k are denoted as $(x_k)_{k \in \mathbb{N}}$.

1.2.2 Statistical formalization

In order to choose the appropriate calibration method, it is necessary to clearly state the assumptions made. This section lists some of the hypotheses that can be found in model calibration applied to mechanical engineering.

The behavior law $\mathcal{F}(\mathbf{t}, \boldsymbol{\theta})$, parameterized by $\boldsymbol{\theta}$ (its spatial variations are neglected), is supposed to mimic the experimental output data $\mathbf{y}$ up to a residual term $\mathbf{r}(\cdot)$ [Kennedy and O'Hagan, 2001, Tarantola, 2005], giving the following decomposition:

$$\mathbf{y} \sim \mathcal{F}(\mathbf{t}, \boldsymbol{\theta}) + \mathbf{r}(\mathbf{t}, \boldsymbol{\theta}) \ . \tag{1.1}$$

Eq. (1.1) can be further detailed into

$$y_j \coloneqq \mathcal{F}(t_j, \boldsymbol{\theta}) + r_j(t_j, \boldsymbol{\theta}) \ \forall j \in [\![1, N]\!] \ , \tag{1.2}$$

with $r_j(t_j, \boldsymbol{\theta})$ the j^{th} coordinate of vector $\mathbf{r}(\mathbf{t}, \boldsymbol{\theta})$. The residuals stand for what cannot be explained by the model to calibrate. Most of the time [Kennedy and O'Hagan, 2001], it is decomposed between experimental noise labeled as $\boldsymbol{\xi}_\mathbf{t} \coloneqq \{\xi_{t_j}\}_{j \in [\![1,N]\!]}$ and model discrepancy (or model inadequacy) $\boldsymbol{\zeta}(\cdot, \cdot)$ as:

$$\mathbf{r}(\mathbf{t}, \boldsymbol{\theta}) = \boldsymbol{\zeta}(\mathbf{t}, \boldsymbol{\theta}) + \boldsymbol{\xi}_\mathbf{t} \ . \tag{1.3}$$

Note that in the following, ξ_{t_j} is simplified into ξ_j. $\boldsymbol{\xi}_\mathbf{t}$ stands for the measurement errors and usually [Arendt et al., 2012], it is distributed according to a Gaussian centered distribution:

$$\xi_j \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \omega_j^2) \ \forall j \in [\![1, N]\!] \ , \tag{1.4}$$

with ω_j the standard-deviation of the j^{th} measurement error and i.i.d standing for independent identically distributed. Note that the i.i.d. hypothesis is a strong assumption that does not always apply. For instance, ignoring correlations between the measurement errors may lead to an overestimation of the noise variance ω_j^2 [Hyslop and Imbens, 2001]. If the ω_js are different for all observations, the noise is said "heteroscedastic". If the standard-deviation is the same for all measurements, then the noise is "homoscedastic" and we can write that $\boldsymbol{\xi}_\mathbf{t} \sim \mathcal{N}(\mathbf{0}_{\mathbb{R}^N}, \omega^2 \mathbf{I}_N)$ with $\mathbf{I}_N$ the identity matrix of $\mathcal{M}_N(\mathbb{R})$ and $\mathbf{0}_{\mathbb{R}^N}$ the zero vector of $\mathbb{R}^N$ [Gallagher and Doherty, 2007]. Heteroscedasticity is sometimes necessary, *e.g.*, when thermal measurements are considered as in [Marquez Costa, 2021].

To explain model bias, let us take a simple example in which the goodness of fit is measured with a mean squared error. As illustrates Figure 1.1, the total error can be decomposed as the sum of the variance term that accounts for the impact of noise on the experimental data and of another term that refers to the difference between the calibrated model and output data that cannot be explained by $\boldsymbol{\xi}$, which is the model discrepancy $\boldsymbol{\zeta}(\cdot)$. In red dashed lines is represented the set of admissible outputs, $\mathcal{Z}$, which corresponds to the image of $\Gamma_{\boldsymbol{\theta}}$ (the research space for the model parameters that implements mathematical and physical constraints) through $\mathcal{F}$ for a given input data $\mathbf{t}$. Let us note $\widehat{\boldsymbol{\theta}}$ the estimated set of model parameters to which is associated model output $\mathcal{F}(\mathbf{t}, \widehat{\boldsymbol{\theta}})$. The quantification of the impact of experimental noise on the model parameters provides a set of model predictions $\mathcal{F}(\mathbf{t}, \boldsymbol{\theta}^{\text{noise}})$, $\mathcal{X}$, represented in blue on Figure 1.1. If the experimental output data $\mathbf{y}$ belongs to $\mathcal{X}$, the noise is sufficient to explain the distance between model output and experimental allowing to conclude that model inadequacy is negligible. On the contrary, if it does not belong to $\mathcal{X}$ as below, one possible explanation is the presence of significant model discrepancy.

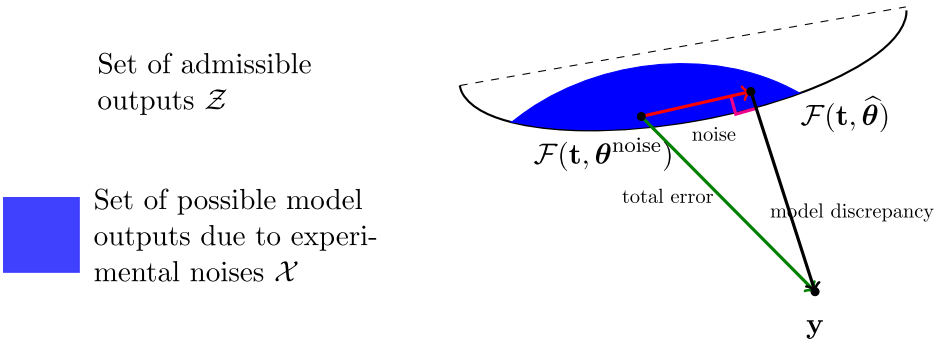


Figure 1.1: Decomposition of the residual term (adapted from [Banks et al., 2014]).

If the model discrepancy is significant, $\boldsymbol{\zeta}(\cdot)$ can be identified with specific techniques for which more details can be found in [Arendt et al., 2012, Banks et al., 2014, Wong et al., 2017]. We hypothesize that, for composites, material variability is the most influential source of uncertainty

before experimental noise, we choose to neglect model discrepancy for the rest of this thesis. Other sources of uncertainty can be included (*e.g.* code uncertainty) but will be also considered as negligible for the remainder of this work. The residual term thus simplifies as follows:

$$\mathbf{r}(\boldsymbol{\theta}) = \mathbf{r} = \boldsymbol{\xi}_{\mathbf{t}} \ . \tag{1.5}$$

This is a strong assumption which may not be true in general and is difficult to (in)validate without considering experimental results. It will be discussed further in the section dedicated to the application of the calibration methods of this chapter. Still, to bring some elements of justification, it is possible to say that behavior laws are supposed to reproduce accurately the experiments but do not account for the intrinsic variability of the material. Furthermore, from a statistical point of view, accounting for model bias significantly increases the complexity of the method and should only be done if it prevents in practice a proper calibration of the available data, which do not seem to be the case in our context as is illustrated for instance in [Laurin, 2005, Marcin et al., 2011].

1.3 Deterministic model calibration

The formalization proposed in Section 1.2 gives an insight of the possible modeling that can be found in model calibration and which can expanded further for specific uses by including other types of uncertainties for example). A usual step in model calibration consists in performing a first deterministic calibration. This solution aims to give a preview of the results, allowing to better tune either the calibration problem (*e.g.* setting more appropriate bounds or the statistical model). It enables to challenge the chosen assumptions: nature of the noise distribution, presence of bias, model formulation, *etc.* It does not always involve an uncertainty quantification phase and only intends to get the best possible prediction which is supposed to be unique, explaining its denomination as "calibration".
This section presents deterministic calibration for both one (1.3.1) or the same repeated experiments (1.3.2) performed on the same material. Some steps of uncertainty quantification are shared with the deterministic calibration presented next such as least squares or likelihood minimization.

1.3.1 Fitting the best curve

The general procedure to calibrate models consists in minimizing over the model parameters a misfit function which quantifies a difference between the model output and the experimental output data [Tarantola, 2005]. This misfit criterion measures how well the model output matches the data. The calibration problem is transformed into the following optimization problem:

$$\min_{\boldsymbol{\theta} \in \Gamma_{\boldsymbol{\theta}}} \mathcal{C}(\boldsymbol{\theta}) \ , \tag{1.6}$$

with $\mathcal{C}(\cdot)$ the fitting criterion to minimize and $\Gamma_{\boldsymbol{\theta}}$ the research space for the model parameters that implements mathematical and physical constraints. Among the usual criteria, one of the most popular is the least squares criterion [Anane et al., 2019, Cornillon and Matzner-Løber, 2007, Janssen and Heuberger, 1995, Tarantola, 2005]. It corresponds to the sum of squares of the differences between the model output and the experimental data and is expressed as follows:

$$\mathcal{C}(\boldsymbol{\theta}) \coloneqq \|\mathbf{y} - \mathcal{F}(\mathbf{t}, \boldsymbol{\theta})\|_2^2 = \sum_{j=1}^{N} (y_j - \mathcal{F}(t_j, \boldsymbol{\theta}))^2 \ , \tag{1.7}$$

with $\|\cdot\|_2$ the L_2 norm. For homoscedastic Gaussian residuals , the least squares provide an estimate of the noise variance $\widehat{\omega}^2$ given by $\widehat{\omega}^2 := \frac{\|\mathbf{y}-\mathcal{F}(\mathbf{t},\widehat{\boldsymbol{\theta}})\|_2^2}{N-d}$, with $\widehat{\boldsymbol{\theta}}$ the calibrated model parameters. The least squares criterion can be transformed into weighted least squares [Kariya and Kurata, 2004, Tarantola, 2005] to balance the weight of the residuals if they appear to be heteroscedastic for example. To improve the behavior of the least squares criterion towards outliers, one can decide to use a robust fitting criterion that ignores the residuals beyond a threshold value [Carroll and Ruppert, 1988]. Other L_p-norms can be used. In particular for $p = 1$, it is called the least absolute deviations criterion [Fernández-Martínez et al., 2011]. Several criteria specific to mechanical problems have also been developed to handle model calibration with field measurements such as FEMU, CEG, *etc.* [Avril et al., 2008].
Another generic criterion can be used to estimate the model parameters: the likelihood [Pintelon and Schoukens, 2012, Wasserman, 2010] denoted $\mathcal{L}(\cdot)$. Given the probabilistic decomposition of the experimental data (see Eq. (1.1) and an assumption on the misfit $\mathbf{r}(\mathbf{t},\boldsymbol{\theta})$, it is defined as the probability density function (PDF) of the residual given its distribution parameters (here ω), $\mathcal{L}(\boldsymbol{\Psi}) := f_{\boldsymbol{\xi}|\omega}(\mathbf{r}(\mathbf{t},\boldsymbol{\theta})|\omega)$, with $f_{\boldsymbol{\xi}|\omega}$ the PDF attached to the statistical modeling of $\mathbf{r}(\mathbf{t},\boldsymbol{\theta})$. In the remainder of this manuscript, $\boldsymbol{\Psi}$ refers to the concatenation of the behavior law parameter $\boldsymbol{\theta}$ and the noise parameters: $\boldsymbol{\Psi} := [\boldsymbol{\theta},\omega]$. For heteroscedastic Gaussian residuals, ω is replaced by $[\omega_1,\dots,\omega_N]$. Under the assumption of i.i.d experiments, the likelihood can be derived as the product of the PDF of the experimental data $\mathbf{y}$ calculated at each experimental time t_j, which gives :

$$\mathcal{L}(\boldsymbol{\Psi}) = \mathcal{L}([\boldsymbol{\theta},\omega]) = f_{\boldsymbol{\xi}|\omega}(\mathbf{r}(\mathbf{t},\boldsymbol{\theta})|\omega) = \prod_{j=1}^{N} f_{\xi_j|\omega}(r_j(t_j,\boldsymbol{\theta})|\omega) \ . \tag{1.8}$$

Contrary to previous criteria, it is maximized rather than minimized to get the maximum likelihood estimate (MLE). The MLE estimate of $\boldsymbol{\Psi}$ is $\widehat{\boldsymbol{\Psi}} = \left[\widehat{\boldsymbol{\theta}},\widehat{\omega}\right]$. In practice, the opposite of the log-likelihood is preferred as it is numerically better conditioned and minimizing is more standard than maximizing in the field of optimization. Indeed, the $f_{\xi_j|\omega}(r_j(t_j,\boldsymbol{\theta})|\omega),\ j \in [\![1,N]\!]$ are below 1. Thus, the global likelihood defined in Eq. (1.8) is very small, close to the machine-epsilon which may lead to accuracy issues. One interest of such a likelihood criterion is to consider a wide range of assumptions on the residual distribution beyond the Gaussian assumption. Note that it corresponds to a least squares criterion if the noise follows a Gaussian distribution. Calibration with least squares is illustrated in [Molimard et al., 2005] for an orthotropic elastic model and in [Chongshuai et al., 2019] for the calibration of a viscoelastic model. Maximum likelihood was used in [Pollak and Palazotto, 2009] for the identification of fatigue curves.

Of course, the choice of the criterion that reflects the chosen assumptions is a key feature. Indeed, it conditions the result and one must ensure it suits the studied problem. Its minimization can be a complicated task and some algorithms were developed specifically, *e.g.* the Levenberg-Marquardt algorithm for least squares problems [Levenberg, 1944, Marquardt, 1963].

To test the validity of the assumptions on the residual model and on the behavior law, model testing can be performed. For instance, to test the normality of residuals, a usual method is to plot the empirical quantiles of the estimated residuals against the theoretical quantiles [Chambers et al., 2018]. The parameterization of the behavior law can be challenged to prevent overfitting. This notion is related to identifiability which can be defined as follows [Walter et al., 1997] :

$$\mathcal{F}(\cdot,\boldsymbol{\theta}_1) = \mathcal{F}(\cdot,\boldsymbol{\theta}_2) \Rightarrow \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 \ . \tag{1.9}$$

In other words, $\boldsymbol{\theta} \mapsto \mathcal{F}(\cdot,\boldsymbol{\theta})$ is injective. In model calibration, overfitting occurs when the data is not informative enough to appropriately estimate all model parameters [Everitt and Skrondal, 2010], that is to say when the model is unidentifiable or when there are too many parameters

in the model. In spite of capturing the global trend, these parameters will be adjusted to capture part of the residual behavior [Anderson and Burnham, 2004]. To detect it, re-sampling methods such as cross-validation or leave-one-out [Allen, 1974] can be used. To prevent it, a common practice is to add to the criteria formulation regularization terms as for example Tikhonov regularization [Tikhonov et al., 1995] or Lasso regression [Tibshirani, 2016]. These terms aim at allowing the calibration of identifiable model parameters and fixing the other parameters to reference values. The non-identifiable can be detected prior to calibration by achieving a sensitivity analysis of the model output with respect to the model parameters with for instance either Morris or Sobol's indices [Morris, 1991, Sobol, 2001], the latter being applied in [Sankararaman et al., 1991]. Such analysis should be done systematically as it can permit to simplify the calibration process by reducing the number of parameters to estimate.
To select the "best" model (combination of the material and residual model), likelihood-based criterion such as Akaike information criterion (AIC) [Akaike, 1974] or Bayesian information criterion (BIC) [Schwarz, 1978] may be employed. Such quantities study at the same time the goodness of fit, the residual model and take into account the total number of parameters, which is here $d+1$ for homoscedastic residuals in the hypothesis of Gaussianity.

1.3.2 Calibration with test repetitions

Mechanical databases are often composed of repetitions of experiments. These repetitions are prescribed by international standards [Defense et al., 1999] to quantify the impact of material variability over the mechanical properties. Repetitions of experiments refer to testing on samples from the same material. In the simplest case, they have the same dimensions, share identical features), the same testing machine is used for all the samples and experimental data is recovered by sensors of the same precision. To take into account several experiments at the same time, deterministic multi-experiment calibration can be used, in particular multi-objective calibration [Madsen, 2003]. It is an extension of single-objective calibration, allowing to handle the calibration of a behavior law with data from several experiments assuming a single value of model parameters for all the experiments. This aims at simultaneously minimizing the fitting criteria associated to each of the experiments that stand for the different objectives. It also provides the trade-offs between the different objectives that appear because of the impact of uncertainties among which one can find measurements errors or intrinsic variability. In the following, an update of the formalization is proposed first. Next, multi-objective calibration is described further.

1.3.2.1 Formalization in the case of test repetitions

Let n be the number of the available repetitions. The output data of these repetitions are denoted $\mathbf{y}_i, \forall i \in [\![1,n]\!]$ associated to their deterministic entries $\mathbf{t}_i$. N_i corresponds the number of observations of the i-th sample. The decomposition of the output data $\mathbf{y}_i$ can be updated from (1.1) to get

$$\mathbf{y}_i \sim \mathcal{F}(\mathbf{t}_i, \boldsymbol{\theta}) + \mathbf{r}(\mathbf{t}_i, \boldsymbol{\theta}) \ . \tag{1.10}$$

Note that in Eq. (1.10), the same set of model parameters is used to estimate all the available experiments. Indeed, the scope of deterministic multi-experiment calibration calibration is to determine the parameters that provide a proper estimation for **all** experiments. The residual still follows the same assumption:

$$\mathbf{r}(\mathbf{t}_i, \boldsymbol{\theta}) = \boldsymbol{\xi}_{\mathbf{t}_i}, i \in [\![1,n]\!] \ , \tag{1.11}$$

where, again, $\boldsymbol{\xi}_{\mathbf{t}_i}$ represents the concatenation of the measurements errors assimilated to a Gaussian white noise with homoscedastic but unknown variance ω_i^2: $\boldsymbol{\xi}_{\mathbf{t}_i} \sim \mathcal{N}(\mathbf{0}_{\mathbb{R}^{N_i}}, \omega_i^2 \mathbf{I}_{N_i})$. Given that the experiments are conducted on the same type of samples, on the same material

and under similar conditions, we assume that the level of noise and measurements errors is comparable for all repetitions. Thus, the parameters of the residual model are the same for all experiments, *i.e.*, $\omega_i = \omega, i \in [\![1, n]\!]$. To each experiment is associated a fitting criterion $\mathcal{C}_i(\cdot), i \in [\![1, n]\!]$. The same set of model parameters $\boldsymbol{\Psi}$ describes the experiment and the multi-objective $\mathcal{C}$ to minimize is:

$$\mathcal{C}(\boldsymbol{\Psi}) \coloneqq [\mathcal{C}_1(\boldsymbol{\Psi}), \dots, \mathcal{C}_n(\boldsymbol{\Psi})] \ . \tag{1.12}$$

The criteria associated to each experiment, $\mathcal{C}_i(\cdot)$, can be for instance the opposite log-likelihoods associated to each experiment, *i.e.* $\mathcal{C}_i(\boldsymbol{\Psi}) = -\ln(\mathcal{L}_i(\boldsymbol{\Psi})) = -\ln(f_{\boldsymbol{\xi}|\omega}(\mathbf{r}(\mathbf{t}_i, \boldsymbol{\theta})|\omega)),\ i \in [\![1, n]\!]$.

1.3.2.2 The multi-objective framework

The simultaneous calibration of different objectives can be carried out in different ways. A simple method consists in considering a weighted sum of the different criteria [Collette and Siarry, 2002]. Another possibility consists in minimizing simultaneously the different objectives, leading to the multi-objective framework [Collette and Siarry, 2002]. The multi-objective framework requires the definition of an order on $\mathbb{R}^n$: the Pareto order denoted $\preccurlyeq$ [Collette and Siarry, 2002, Jahan and Edwards, 2013]. Given two set of model parameters $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$, $\boldsymbol{\Psi}_1 \preccurlyeq \boldsymbol{\Psi}_2$ if

- $\exists I \subset [\![1, n]\!], I \neq \varnothing | \forall i \in I, \mathcal{C}_i(\boldsymbol{\Psi}_1) < \mathcal{C}_i(\boldsymbol{\Psi}_2)$,
- $\forall i \in [\![1, n]\!] \backslash I, \mathcal{C}_i(\boldsymbol{\Psi}_1) \leqslant \mathcal{C}_i(\boldsymbol{\Psi}_2)$.

In other words, $\boldsymbol{\Psi}_1$ is better than $\boldsymbol{\Psi}_2$ for at least one objective and $\boldsymbol{\Psi}_1$ is at least as good as $\boldsymbol{\Psi}_2$ for the others. In Figure 1.2, the Pareto-dominance is illustrated: $C \preccurlyeq D \preccurlyeq F$, $A \preccurlyeq E$

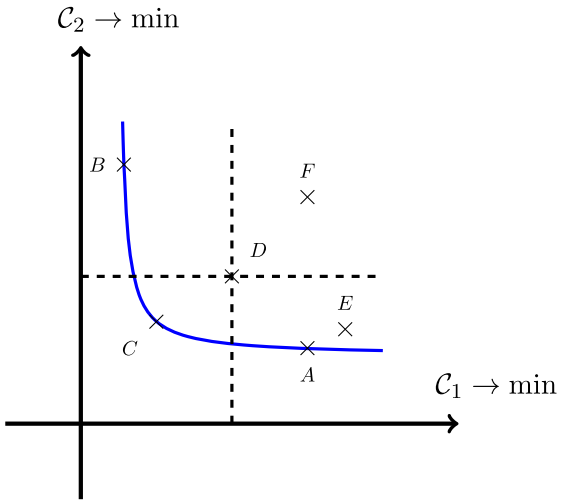


Figure 1.2: Illustration of Pareto-dominance in $\mathbb{R}^2$.

and $B \preccurlyeq F$. A, B and C are "non-dominated" points and belong to $\mathcal{P}$ called the Pareto front, represented in blue in Figure 1.2. It contains the best trade-offs between the different criteria associated to the experiments [Efstratiadis and Koutsoyiannis, 2010]. The Pareto order is a weak order, all $\boldsymbol{\Psi}$s are not comparable and consequently, there is no unique result to the minimization problem. With regard to the fitting criterion, model parameters in $\mathcal{P}$ exhibit similar goodness of fit [Efstratiadis and Koutsoyiannis, 2010]. There are other orders in $\mathbb{R}^n$ that may be either derived from Pareto dominance (extreme Pareto optimality) or total orders (lexicographic order) [Ehrgott, 1999]. A more complete list can be found in [Efstratiadis and Koutsoyiannis, 2010].

To perform the multi-objective minimization, the multi-objective problem can be transformed into a single-objective problem [Van Veldhuizen, 1999] considering a weighted sum of the objectives (that enables to determine the convex part of $\mathcal{P}$). Genetic algorithms can also be used to optimize simultaneously all criteria [Efstratiadis and Koutsoyiannis, 2010]. Multi-objective calibration is illustrated in [Rajib et al., 2016] for the identification of an hydrological model.

The Pareto front is said non-degenerated if it is not reduced to a single set of model parameters, meaning that trade-offs between the different objectives are made. The presence of trade-offs indicates the presence of uncertainties [Rosbjerg and Madsen, 2006]. To explain this point, let us imagine a situation with no uncertainties of any type and no bias in the model. In such a case, there would not be any inadequacy between the criteria as all experiments give the same information and the Pareto front would be reduced to a single point. Yet, in the presence of uncertainties, small incompatibilities appear between the objectives leading to the apparition of trade-offs. What makes it difficult to directly interpret Pareto results is that the impact of the different uncertainties (measurement errors, model bias, *etc.*) is mixed and it is difficult to extract the consequences of a specific uncertainty. Finally, it cannot handle the calibration of many repetitions because an increasing proportion of model parameters belongs to Pareto front as the number of objectives increases. In practice above five competing objectives, one expects that most parameter settings will belong to the Pareto front [Coello, 2005].

This section has presented deterministic model calibration, with a focus on the case where repetitions of experiments are available. Note that this estimation does not account for the impact of uncertainties on the model parameters that can be estimated in either the frequentist framework (*cf.* the next section) or in a Bayesian fashion (*cf.* Section 1.5).

1.4 Uncertainty quantification with the frequentist approach

The previous section aimed at identifying a global trend that can be used as a first approximation characterized by a unique set of model parameters for one experiment (and if available the corresponding repetitions). However, given that it usually remains a residual error for the estimated set of model parameters, relying on single set of model parameters is not satisfying as there can be different sets of model parameters that can explain an experiment with the same degree of precision. Quantifying the impact of uncertainties on the model parameters allows to account for the consequences of measurement errors or model inadequacy and dispersion in the material used in the experiments. This impact can be illustrated by confidence intervals or regions. Intervals describe the variability coordinates by coordinates when regions characterize both the variability of each model parameters and the correlations between the estimates of parameters. This part details two main methods to perform this estimation within the frequentist framework. Note that these methods apply to the case where a single experiment is calibrated.

1.4.1 Uncertainty quantification with the asymptotic theory

The estimation of parameter uncertainty can be carried out with the asymptotic theory [Antoniadis and Carmona, 1992, Banks et al., 2014, Gallagher and Doherty, 2007] which provides estimations of trust regions at a desired confidence level. This theory demands to meet two main requirements. First, the number of observations should be large enough as it is based on asymptotic results, that is to say when $N \to +\infty$. Second, the residuals $\mathbf{r}$ should be Gaussian distributed. For a least squares criterion, the next result applies:

$$\boldsymbol{\theta} \sim \mathcal{N}(\widehat{\boldsymbol{\theta}}, \widehat{\omega}^2 \mathcal{J}_N(\widehat{\boldsymbol{\theta}})^{-1}) \text{ with } \mathcal{J}_N(\widehat{\boldsymbol{\theta}}) \coloneqq \sum_{j=1}^{N} \nabla_{\boldsymbol{\theta}} \mathcal{F}(t_j, \widehat{\boldsymbol{\theta}}) \nabla_{\boldsymbol{\theta}} \mathcal{F}(t_j, \widehat{\boldsymbol{\theta}})^\top \ . \tag{1.13}$$

In Eq. (1.13), $\widehat{\boldsymbol{\theta}}$ is the least squares estimate and $\mathcal{J}_N(\widehat{\boldsymbol{\theta}})$ refers to the Gauss-Newton approximation of the Hessian matrix of a least squares based criterion [Antoniadis and Carmona, 1992, Banks et al., 2014]. For each model parameter θ_l, the confidence interval at level $1-\alpha$ is denoted $I^l_{1-\alpha}$ and is given by:

$$I^l_{1-\alpha} \coloneqq \{\theta_l \in \mathbb{R} |\ |\theta_l - \widehat{\theta}_l| \leqslant \sqrt{\widehat{\omega}^2 \mathcal{J}_N(\widehat{\boldsymbol{\theta}})_{ll}} T^{-1}_{N-d}(1-\frac{\alpha}{2})\} \ \forall l \in [\![1,d]\!] \ , \tag{1.14}$$

with T^{-1}_{N-d} the quantile of the Student distribution with $N-d$ degrees of freedom and $\mathcal{J}_N(\widehat{\boldsymbol{\theta}})_{ll}$ the l^{th} diagonal coefficient of $\mathcal{J}_N(\widehat{\boldsymbol{\theta}})$. The confidence region at level $1-\alpha$ of model parameter is given by:

$$R_{1-\alpha} = \{\boldsymbol{\theta} \in \mathbb{R}^d \ \mid \ (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^\top \mathcal{J}_N(\widehat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) \leqslant \widehat{\omega}^2 d F^{-1}_{d,N-d}(1-\alpha)\} \ , \tag{1.15}$$

with $F^{-1}_{d,N-d}$ the quantile of the Fisher distribution with degrees of freedom $d, N-d$. The confidence sets defined by Eqs. (1.14) and (1.15) are based on a linearization of the model $\mathcal{F}(\mathbf{t},\cdot)$ around $\widehat{\boldsymbol{\theta}}$. The exact formulas without approximations of the non-linear confidence sets can be found in [Antoniadis and Carmona, 1992]. More generally, under regularity conditions of the likelihood, the MLE estimator exhibit a similar asymptotic behavior [Ly et al., 2007, Wasserman, 2010]:

$$\boldsymbol{\Psi} \sim \mathcal{N}(\widehat{\boldsymbol{\Psi}}, \frac{1}{N}\mathcal{I}_N(\widehat{\boldsymbol{\Psi}})^{-1}) \text{ with } \mathcal{I}_N(\boldsymbol{\Psi}) \coloneqq \mathbb{E}_{\boldsymbol{\Psi}}\left[\frac{\partial^2 \ln(\mathcal{L}(\boldsymbol{\Psi}))}{\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}^\top}\right] \tag{1.16}$$

the Fisher Information Matrix (FIM). Note that with a Gaussian likelihood, the FIM corresponds to the Hessian of the log-likelihood [Kunstner et al., 2019, Martens, 2020]. Example of applications of these formulas can be found in [Cokelaer, 2008, Guida and Pulcini, 2009, Lu et al., 1997]. When the residuals do not follow a Gaussian distribution, confidence sets based on the asymptotic behavior of the MLE can be also derived thanks to other methods such as the likelihood ratio [Nelson, 2004] among others. Let us split $\boldsymbol{\Psi}$ into two subsets $\boldsymbol{\Psi}_1 \in \mathbb{R}^{d_1}$ and $\boldsymbol{\Psi}_2 \in \mathbb{R}^{d+s-d_1}$ and define the restricted likelihood $\mathcal{L}_1$ as $\mathcal{L}_1(\boldsymbol{\Psi}_1) \coloneqq \mathcal{L}(\boldsymbol{\Psi}_1, \widehat{\boldsymbol{\Psi}}_2)$, that is to say the likelihood function with $\boldsymbol{\Psi}_2$ fixed to the MLE estimate. The trust region at level $1-\alpha$ is defined by:

$$R_{1-\alpha} = \{\boldsymbol{\Psi}_1 \in \mathbb{R}^{d_1} | (\ln(\mathcal{L}_1(\boldsymbol{\Psi}_1)) - \ln(\mathcal{L}(\widehat{\boldsymbol{\Psi}}))) \leqslant (\chi^2_{d_1})^{-1}(1-\alpha)\} \ , \tag{1.17}$$

with $(\chi^2_{d_1})^{-1}(1-\alpha)$ the quantile of the chi-squared distribution with d_1 degrees of freedom. Under the assumption of a Gaussian distributed residual, $\boldsymbol{\Psi} = [\boldsymbol{\theta}, \omega]$ (so $\boldsymbol{\Psi}_1 = \boldsymbol{\theta}$ and $\boldsymbol{\Psi}_2 = \omega$). Note that the FIM or the model gradient $\mathcal{J}_N$ (defined in Eq. (1.13)) can be also used for other purporses. For instance, they can be used to detect unidentifiable parameters once model calibration is completed by analyzing their rank and eigenvalues [Cintrón-Arias et al., 2009, Cobelli and DiStefano, 1980]. This property has been used to design an iterative procedure proposed in [Anane et al., 2019]. Following a first calibration, one of these two matrices is computed and its eigenvalues analyzed. Then, the combination of parameters corresponding to the null eigenvalues are set to a nominal values. This scheme is repeated until all eigenvalues are non-zero. A similar analysis can be done with the eigen-vectors [Molimard and Le Riche, 2003]. The model gradient $\mathcal{J}_N$ can be also be used to strengthen the identification of the model parameters as proposed in [Meraghni et al., 2014]. Taking into account such information allows to better adapt the minimization algorithm to the local variations of the fitting criterion.

In spite of its efficiency, this method to derive confidence intervals is limited by several aspects. First, it is based on an asymptotic behavior. Therefore, its significance when the number of observations is low may be questioned. Second, it is based on Gaussian distributed residuals, an assumption that should be checked. Finally, from a numerical point of view, the method can

fail [Hu et al., 2011]. Indeed, it demands to compute derivatives whose precision can prevent the FIM from being invertible [Martens, 2020]. In that case, the use of Penrose pseudo-inverse may be considered [Barata and Hussein, 2011].

1.4.2 Uncertainty quantification with bootstrap-based techniques

Because of non-Gaussian residuals or a fitting criterion with poor regularity properties, the asymptotic theory may not work. To estimate confidence intervals and regions, the bootstrap can be applied [Efron and Tibshirani, 1994]. Its main advantage is that it applies without the need of major assumptions. The general principle consists in providing an empirical sampling from the distribution of $\boldsymbol{\Psi}$ (N_B samples here). Given that the calibration method (least squares, maximum likelihood) only provides one sample of the model parameters given a set of experimental observations $\mathbf{y}$, to get multiple (i.i.d.) samples, the experimental data is resampled to form the bootstrapped output data $\mathbf{y}_{\text{boot},k} \coloneqq (y_{\text{boot},k,j})_{j\in[\![1,N]\!]}$ for $k \in [\![1, N_B]\!]$. For each of these bootstrapped datasets, the behavior law $\mathcal{F}(\mathbf{t}, \cdot)$ is calibrated to find bootstrap samples $\{\boldsymbol{\Psi}_{\text{boot},k}\}_{k\in[\![1,N_B]\!]}$ [Banks et al., 2014, Efron and Tibshirani, 1994, Wu et al., 2018]. The convenient ways to bootstrap data are discussed later. The bootstrap estimator is asymptotically consistent [Mammen, 1992]: the empirical distribution of the bootstrapp samples converges to the distribution of $\boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is the exact distribution of model parameters. More precise definitions of consistency and theorems to ensure it can be found in [Mammen, 1992]. If it applies, with the empirical samples of $\{\boldsymbol{\Psi}_{\text{boot},k}\}_{k\in[\![1,N_B]\!]}$, it is possible to estimate variances, correlations, confidence regions and intervals, *etc.* To construct confidence intervals, two solutions can be chosen [Carpenter and Bithell, 2000]. The first consists in ordering the bootstrap samples. The $(1-\alpha)100\%$ confidence interval is defined by the $\alpha 100^{\text{th}}\%$ and $(1-\alpha)100^{\text{th}}\%$ quantiles of the empirical bootstrap distribution. The quantiles are given by the $(k+1)\alpha^{\text{th}}$ and $(k+1)(1-\alpha)^{\text{th}}$ elements of the ordered bootstrap samples. Another possibility is to approximate the empirical distribution of the bootstrap samples with a Gaussian distribution giving, for the confidence interval $I^l_{1-\alpha}$ at level $1-\alpha$ of each component of vector $\boldsymbol{\Psi}$ labeled $[\Psi_l]_{l\in[\![1,d+1]\!]}$:

$$I^l_{1-\alpha} = [\Psi_{l,\text{med}} - q_{1-\frac{\alpha}{2}}\text{sd}(\Psi_{l,\text{boot}}), \Psi_{l,\text{med}} + q_{1-\frac{\alpha}{2}}\text{sd}(\Psi_{l,\text{boot}})] \ , \tag{1.18}$$

with $\boldsymbol{\Psi}_{\text{med}}$ the median of the bootstrap samples and $\text{sd}(\Psi_{l,\text{boot}})$ the empirical standard deviation of the bootstrap samples defined as:

$$\text{sd}(\Psi_{l,\text{boot}}) \coloneqq \sqrt{\frac{1}{N_B}\sum_{k=1}^{N_B}(\Psi_{l,\text{med}} - \Psi_{l,\text{boot}})^2} \ , \tag{1.19}$$

Bootstrap is applied to quantify uncertainties on parameters of a Finite Element model [Gibanica et al., 2016], on a fatigue model in [Heiermann et al., 2005] and on composite beam model in [Hu et al., 2011].

The differences between the types of bootstrap lie in the way data is resampled. A first kind of bootstrap is the parametric bootstrap [Horowitz, 1996]. It relies on the assumption made on the residuals parameterized by ω. To apply it, the first step consists in calibrating the model parameters and the noise parameters, in other words $\widehat{\boldsymbol{\Psi}}$. Once this step is completed, for all $k \in [\![1, N_B]\!]$, a noise profile $\boldsymbol{\xi}_{\text{boot},k}$ is drawn from the estimated distribution of the residuals and added to the calibrated model $\mathcal{F}(\mathbf{t}, \widehat{\boldsymbol{\theta}})$ to form the bootstrapped data $\mathbf{y}_{\text{boot},k}$. The estimation of bootstrap samples can be then carried out.

Parametric bootstrap has one major drawback: it relies on the assumption of the residual model (presence of bias, nature of measurements errors, *etc.*) and the results can be flawed

if the hypotheses are not met. In order not to depend on these assumptions, non-parametric versions have been developed [Wu et al., 2018]. Another way to resample experimental data $\mathbf{y}$ consists in randomly permuting the observations. However, this process has a major drawback. For instance, an observation made at the end of the experiment can be moved to the start and conversely. This may lead to the creation of data that does not represent physical observations, especially when the model is time dependent and even more if the model explicitly depends on the time variable. Consequently, rather than the data, the residuals can be permuted [Liu, 1988]. In practice, following a first calibration, the optimal residuals $\widehat{\mathbf{r}}$ are determined by $\widehat{\mathbf{r}} = \mathbf{y} - \mathcal{F}(\mathbf{t}, \widehat{\boldsymbol{\theta}})$ and their components are then randomly chosen. For all steps $k \in [\![1, N_B]\!]$, N discrete uniform numbers between 1 and N are sampled: $\{U_{j,k}\}_{j \in [\![1,N]\!]}$; $U_{j,k} \sim \mathcal{U}([\![1, N]\!])$. These numbers represents the indices of the bootstrapped residuals $\mathbf{r}_{\text{boot},k}$ defined as $r_{\text{boot},j,k} := \widehat{r}_{U_{j,k}}, j \in [\![1, N]\!]$. The bootstrapped data is obtained as $\mathbf{y}_{\text{boot},k} = \mathbf{y} + \mathbf{r}_{\text{boot},k}$. Note that if the residuals are heteroscedastic, it may be necessary to rescale them or to swap only those with comparable variance. After data generation, the estimation of bootstrap samples can be then achieved.

To reach statistical significance of the empirical distribution of the bootstrap samples, whatever variant is chosen, it is necessary to achieve enough bootstrap iterations, that is to say to have N_B high enough. This is the main drawback of this approach, especially if the behavior law requires important numerical facilities. This procedure can be parallelized on several machines but remains computationally intensive as a Monte-Carlo method.

The usual frequentist methods for uncertainty quantification have been described. Note that both the asymptotic theory and the bootstrap require to perform a first deterministic calibration. Yet, uncertainty quantification can be carried out in a single step in a Bayesian fashion, which is presented next.

1.5 Bayesian method for model calibration

1.5.1 Introducing Bayesian inference

It is possible to consider different points of view on the estimation of model parameters: among them the frequentist and Bayesian methods [Wasserman, 2010]. Section 1.4 dealt with the frequentist approaches devoted to model calibration in the presence of uncertainty. The fundamental difference between the two approaches is that, in the frequentist paradigm, the model parameters are considered as fixed but unknown (they are intrinsically deterministic). This lack of knowledge is illustrated *via* approximated confidence sets that asymptotically contain the "true" value of the model parameters. In a Bayesian vision, model parameters are considered as random variables. The objective of Bayesian inference is to update the prior knowledge on the model parameters by learning the information in the data. From this information, it is possible to infer their posterior probability distribution that can be used to obtain credible interval or carry out propagations, *etc.*

Before presenting the main features of Bayesian inference, we make a general assumption: all considered distributions will be supposed to have a probability density function (PDF), noted f in the following. The objective of Bayesian inference is to determine the distribution of $\boldsymbol{\Psi}$ given the experimental data, which is called the posterior density: $f(\boldsymbol{\Psi}|\mathbf{y})$ [Box and Tiao, 2011, Kennedy and O'Hagan, 2001, Wasserman, 2010]. It represents the impact of uncertainties on model parameters. The approach relies on the definition of a prior distribution illustrating the previous knowledge $f(\boldsymbol{\Psi})$ (detailed further in Section 1.5.2). Given the Bayes theorem, the posterior can be computed if one knows the density of the data given the model parameters,

$f_{\boldsymbol{\xi}|\omega}(\mathbf{r}(\mathbf{t},\boldsymbol{\theta})|\omega)$ (the likelihood introduced in 1.3.1):

$$f(\boldsymbol{\Psi}|\mathbf{y}) = \frac{f(\boldsymbol{\Psi}) f_{\boldsymbol{\xi}|\omega}(\mathbf{r}(\mathbf{t},\boldsymbol{\theta})|\omega)}{f(\mathbf{y})} \ , \tag{1.20}$$

with $f(\mathbf{y})$ a normalization constant that does not depend on the parameters to estimate $\boldsymbol{\Psi}$ and can be expressed as:

$$f(\mathbf{y}) \coloneqq \int_{\mathbb{R}^{d+1}} f(\boldsymbol{\Psi}) f_{\boldsymbol{\xi}|\omega}(\mathbf{r}(\mathbf{t},\boldsymbol{\theta})|\omega) \mathrm{d}\boldsymbol{\Psi} \ , \tag{1.21}$$

with d the number of parameters of the behavior law which is also the number of parameters of the residual model plus 1 parameter, ω, for the noise. Here, model calibration and uncertainty quantification are carried out together. Bayesian inference has been applied in [Rappel et al., 2019a] to the identification of a viscoelastic model, in [Gogu et al., 2013] to the identification of a Finite Element model and in [Paranjape et al., 2021] to the identification of a superelastic model.

1.5.2 Prior density

The prior density lies at the core of the Bayesian approach. Indeed, its choice partly determines the posterior as for two different priors, the posterior will be different [Berger, 1985]. Furthermore, it represents the "degree of belief" introduced in the Bayesian inference. Second, from a computational point of view, the prior is necessary to evaluate the posterior. Two types of priors can be selected: subjective or non-informative priors.

Contrary to their non-informative counterpart, subjective priors intend to bring a significant amount of knowledge available before calibration in the process. For instance, this knowledge corresponds to physical and mathematical constraints on the considered parameters but also to information given by material producers or specialists [Gogu et al., 2013]. Subjective priors may also be inferred from experimental data complementary to the one involved in the calibration process. This is known as Empirical Bayes [Berger, 1985]. Subjective prior can also be chosen to ease calculations. Indeed, for some likelihood functions, a suitable prior choice enables an analytical derivation of the posterior distribution. When the prior and posterior belong to the same family of probability distributions, the prior is called conjugate prior with respect to the likelihood. For some likelihoods function, the exponential family is a conjugate prior. In model calibration, with a linear model and Gaussian distributed residuals, a Gaussian prior gives a Gaussian posterior that can be analytically derived [Bouloré, 2019, Rappel et al., 2019a].

One can also choose not to influence the calibration process with excessively informative priors [Wasserman, 2010]. These priors are called non-informative priors. Among them can be found Jeffreys priors [Jeffreys, 1946] or maximum entropy priors [Rosenkrantz, 1989]. Maximum entropy priors [Rosenkrantz, 1989] aim to be the least informative possible given a set of constraints. Indeed, the entropy, defined as $-\mathbb{E}(\ln(X))$ for a positive random variable X [Shannon, 1948], measures the amount of information of a random variable. The constraints stand for support, integrability, mean, variance, *etc.* [Cover and Thomas, 2006]. If parameters are bounded, the corresponding prior is a uniform prior (which corresponds to the intuition) and if the support is not bounded, given means and variances, it is Gaussian. Yet, these priors are not fully satisfactory as they depend on the chosen parameterization as the uniform prior. Jeffreys developed a class of priors proportional to the square root of the determinant of the FIM [Jeffreys, 1946] that are independent of the chosen parameterization. Yet, they are improper priors as their corresponding density does not integrate to 1, which makes them difficult to use.

1.5.3 Posterior sampling

Except in very specific cases, the exact posterior cannot be derived, especially because of the denominator of Bayes' rule defined in Eq. (1.21) whose computation is often intractable. One solution consists in getting samples from the posterior distribution. To achieve this task, it is possible to use "Monte-Carlo Markov chain" (MCMC) algorithms [Robert and Casella, 2005]. Among the usual sampling algorithms, one can find the Metropolis-Hastings (M-H) [Hastings, 1970, Metropolis et al., 1953] or the Gibbs sampler [Tanner, 1998]. These algorithms build a Markov Chain (a random walk) within the parameters space and either accept or reject with their own procedure the proposed samples. Under general assumptions, the empirical distribution of the samples converges to the exact posterior distribution.

The M-H algorithm is based on acceptation-rejection methods [Devroye, 1986, Robert and Casella, 2005] and generalizes the sampling of any distribution whose density can be computed up to a normalization constant. Bayesian calibration fits with this situation as the posterior density can be computed up to a normalization constant following Eq. (1.20). The algorithm works as follows. For a given set of initial model parameters (the MLE for instance) $\boldsymbol{\Psi}_0$, the algorithm draws a candidate sample $\boldsymbol{\Psi}_*$ from a proposal density $q(\boldsymbol{\Psi}_*|\boldsymbol{\Psi}_0)$. The role of this density is only to generate samples to explore the parameters space. In practice, it is often either a Gaussian or a uniform distribution [Rappel et al., 2019b]. To accelerate the convergence of the algorithm, it can be given a specific shape [Gelman et al., 1996]. The rest of the algorithm is based on an acceptation-rejection method: a ratio is computed, here $r(\boldsymbol{\Psi}_*, \boldsymbol{\Psi}_0) := \frac{f(\boldsymbol{\Psi}_*|\mathbf{y})q(\boldsymbol{\Psi}_0|\boldsymbol{\Psi}_*)}{f(\boldsymbol{\Psi}_0|\mathbf{y})q(\boldsymbol{\Psi}_*|\boldsymbol{\Psi}_0)}$. Next, a uniform number u is drawn from a uniform distribution between 0 and 1. If the ratio is greater than u, the sample $\boldsymbol{\Psi}_*$ is accepted and becomes $\boldsymbol{\Psi}_1$, otherwise it is rejected. Note that if the ratio is greater than 1, then the $\boldsymbol{\Psi}_*$ is always accepted. In fact, with a symmetric proposal, this ratio indicates if the proposed sample is more probable regarding the posterior distribution than the previous sample. The full algorithm is described in 1.

Algorithm 1 Metropolis-Hastings algorithm

Require: initial sample $\boldsymbol{\Psi}_0$, number of iteration N_{it}
for $k = 1, \ldots, N_{\text{it}}$ **do**
  Generate sample $\boldsymbol{\Psi}_* \sim q(\boldsymbol{\Psi}_*|\boldsymbol{\Psi}_k)$
  Compute $r(\boldsymbol{\Psi}_*, \boldsymbol{\Psi}_k)$
  Draw $u \sim \mathcal{U}[0, 1]$
  if $r(\boldsymbol{\Psi}_*, \boldsymbol{\Psi}_k) > u$ **then**
    Accept the proposed sample and $\boldsymbol{\Psi}_{k+1} = \boldsymbol{\Psi}_*$
  else
    Reject the proposed sample and $\boldsymbol{\Psi}_{k+1} = \boldsymbol{\Psi}_k$
  end if
end for

Once all steps are completed, the first samples generated are usually discarded to prevent the potential bias induced by the choice of $\boldsymbol{\Psi}_0$. This is called the "burn-in" [Andrieu et al., 2003, Rappel et al., 2019b]. In theory, the generated samples should be independent. However, given the generation procedure, the samples are correlated. To address this issue, a common practice is the "thinning" which consists in sub-sampling the selected samples to lower the correlation between the different samples [Gelman et al., 2014, Link and Eaton, 2012].

Numerous other sampling algorithms can be found. Among them, one can find variants of the M-H algorithm with the Adaptative Proposal Metropolis Hastings presented in [Haario et al., 1999], or other schemes such as the Langevin dynamics [Robert and Casella, 2005]. The difference between the latter and the M-H algorithm resides in the construction of the random

walk. Another popular sampling algorithm is the Gibbs sampler [Tanner, 1998]. This algorithm samples parameters coordinate by coordinate with an acceptance rate of 1. It is mainly used if it is easier to sample the marginal posterior densities rather than the joint posterior distribution.

The convergence of MCMC methods is a complex issue. Convergence can be defined by several criteria. Among them, one can find the convergence of the empirical distribution to a stationary distribution and the convergence to i.i.d. samples [Robert and Casella, 2005]. Generally, there is not clear answer to this problem and it is necessary to study multiple items, some of which are defined in [Sinharay, 2003].
Aside from these usual techniques, many other calibration methods can be implemented. For instance, techniques based on Kalman filtering [Kalman, 1960] are particulary suited for real-time assimilation as applied in [Marchand et al., 2016] in structural mechanics.

The usual (frequentist and Bayesian) methods to quantify the uncertainty on models parameters, (frequentist and Bayesian) have been presented. However, given the duration of mechanical simulation (*e.g.,* with Finite element Analysis) it may be necessary to implement strategies that reduce the computational costs.

1.6 Speeding up the calibration process

All previous approaches usually demand a large number of model evaluations, either to achieve the minimization of the chosen fitting criterion or to guarantee a minimum number of samples with both Bayesian and bootstrap procedures. Given the duration of model evaluations, it is necessary to alleviate computational costs. To accelerate model calibration, two approaches can be followed. The first is to reduce the cost of model evaluations thanks to surrogate modeling. Another strategy is to reduce the number of model parameters to calibrate by fixing unidentifiable parameters.

To alleviate the cost of model evaluations, surrogate models have been proposed [Forrester et al., 2008]. Surrogate models are approximations of the true model that are much faster to evaluate, allowing for proper calibration. Among the different types of surrogates, one can find Gaussian process proposed in [Martin and Simpson, 2005] and applied in [Hu et al., 2017], radial basis functions applied in [Pacheco et al., 2016], support vector machines [Forrester et al., 2008], polynomial chaos proposed in [Ghanem et al., 1991] and applied in [Janouchová et al., 2021], or artificial neural networks described in [Jain et al., 2005] and applied in [Heiermann et al., 2005]. The dimension of the problem can be reduced by projecting the model on specific bases. Such techniques are derived from the Karhunen–Loève expansion [Loeve, 1978] and include the Proper Order Decomposition applied in [Gogu et al., 2013]. Note that when surrogate models are used, it is necessary to account for the approximation error in the statistical formalization by introducing another error term in the residual defined in Eq. (1.3).

The above literature review has presented tools that are mainly meant to achieve uncertainty quantification of (expensive) behavior law parameters. These methods will be now applied to a mechanical problem to study their performances, in particular for the estimation of material variability.

1.7 Application to the calibration of a uniaxial behavior law

The previous sections that consisted in a literature review now over have presented standard model calibration techniques in the presence of experimental uncertainties. Such methods can handle experimental noise and to a lesser extent, model inadequacy. This part aims to show their limitations in the estimation of the intrinsic variability of the mechanical properties, which is

known to be critical for composites material. It arises by comparing repetitions of experiments. After presenting the experiments, the significance of material variability will be shown. Then, an attempt to characterize the material variability such as it is seen by the methods described until now will be conducted. Their limitations will be discussed. The behavior law used to mimic these experiments, the ONERA Damage Model (ODM) [Maire et al., 1997, Marcin et al., 2011, Ben Ramdane, 2014] is described the next section.

1.7.1 Presentation of the mechanical test-case

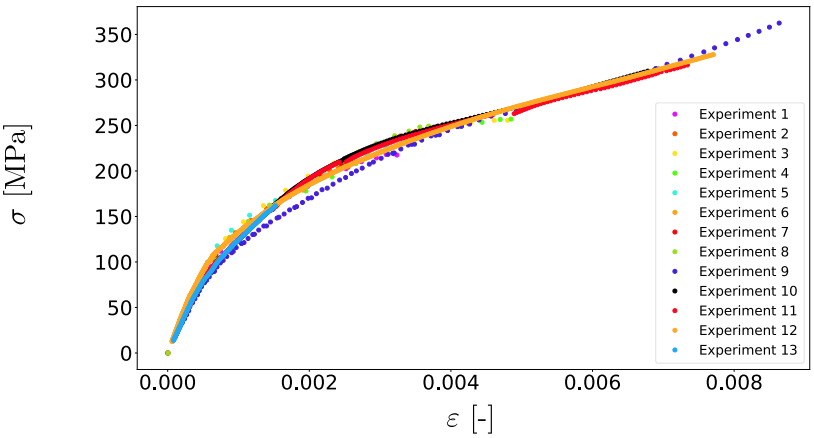


Figure 1.3: 13 experimental monotonic tensile strain-stress curves on Ceramic Matrix Composite material [Naslain, 2004, Viricelle et al., 2001].

Figure 1.3 depicts 13 tensile tests carried out on the CERASEP A400 woven composite material [Naslain, 2004, Viricelle et al., 2001]. The tensile test are performed with monotonic loadings until failure. These curves exhibit two behaviors: an elastic linear behavior up to a threshold after which the stiffness of the material decreases non-linearly because of damage. The transition between the linear and non-linear regimes is subject to a significant variability, even if the curves tend to overlap when damage increases, for strains superior to 0.4 %. At the end of the experiment, before failure, the damage seems to reach saturation, and the behavior of the experiments can be compared to a linear elastic behavior with damaged stiffness. Notice also that the strain-stress curves have different numbers of points (from 20 to 700 points).
The model of interest, the simplified uni-axial ODM model. It relates the stress (in MPa) labeled σ to the (dimensionless) strain labeled ε. The strain-stress relation is

$$\sigma(t) = E^{\text{eff}}(d_s(t))\varepsilon(t) \ , \tag{1.22}$$

with E_0 the Young's modulus (in GPa), E^{eff} the effective Young's modulus which takes into account the loss in stiffness of the material caused by the damage d_s. The effective Young's modulus reads as

$$E^{\text{eff}}(d_s(t)) = \frac{E_0}{1+\eta(t)d_s(t)} \ . \tag{1.23}$$

The damage desactivation index, η, represents the fact that damage does not impact the stiffness for compressive loads. It is given by $\eta(t) = h(\varepsilon(t))$ with h the Heaviside function. The damage d_s is computed from the thermodynamical force y. The damage equations are the following:

$$\begin{cases} y(t) = \dfrac{1}{2}E_0\langle\varepsilon(t)\rangle_+^2 \ , & \text{(1.24a)} \\ g_s(y(t)) = \dfrac{\sqrt{y_{max}(t)} - \sqrt{y_{0s}}}{\sqrt{y_{cs}}} \text{ with } y_{max}(t) = \sup\limits_{\tau\in[0,t]} y(\tau) \ , & \text{(1.24b)} \\ d_s(t) = d_c(1-\exp(-\langle g_s(y(t))\rangle_+^p)) \ , & \text{(1.24c)} \end{cases}$$

where $\langle x \rangle_+$ stands for the positive part of x (*i.e.* 0 if $x < 0$ and x if $x > 0$). Eq. (1.24a) defines the thermodynamical driving force. The positive part means that damage is created only under tensile stress. In Eq. (1.24b), y_{0s} is the damage evolution celerity (in MPa) and y_{cs} is the damage evolution celerity (in MPa). In Eq. (1.24c), d_c is the damage saturation (without units) and p a shape parameter called the damage evolution exponent (without units). It can describe both linear and non-linear behavior. The input data refer to the strain which will be supposed as perfectly known and the (noisy) output to the stress profile. The Young's modulus is the only parameter that controls the elastic linear part. The damage threshold indicates when the damage starts (equivalently the non-linear behavior). To this damage threshold can be associated a strain damage threshold given by $\varepsilon_{0s} = \sqrt{\frac{2y_{0s}}{E_0}}$, represented by a green line in Figure 1.4. For pristine materials and deformations smaller than ε_{0s}, the damage is equal to 0 and the material exhibits a linear behavior. The damage evolution celerity controls the speed at which damage develops: high y_{cs}s mean slower damage and conversely. The damage evolution exponent monitors the smoothness of the transition between the linear and nonlinear regime. This parameter is heavily correlated to the other parameters in this model. Consequently, for the sake of simplicity, it is set to 1.2, a usual value for this kind of material. Note that in this case, the damage threshold is difficult to estimate and is usually determined from acoustic emission data that is not available here [1]. The damage saturation is the maximum possible damage before failure and can be understood as the highest possible loss of stiffness for the considered experiment. The asymptotic slope refers to $\frac{E_0}{1+d_c}$ as represented in green in Figure 1.4. Finally, 4 parameters have to be calibrated, all of them have to be non-negative. Their signification is illustrated in Figure 1.4 and they are summarized in Table 1.1.

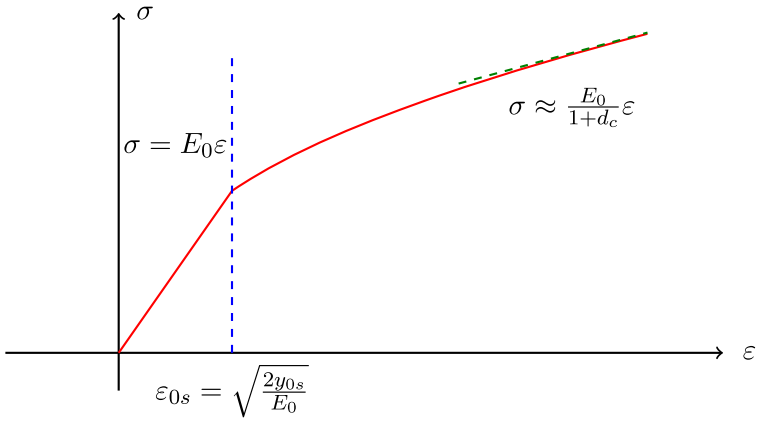


Figure 1.4: Interpretation of the ODM model parameters

Table 1.1: ODM Model parameters to calibrate.

Young's modulus	damage threshold	damage evolution celerity	damage saturation	damage evolution exponent
E_0 [GPa]	y_{0s} [kPa]	y_{cs} [MPa]	d_c [-]	$p = 1.2$ [-]

1.7.2 Deterministic calibration of the experiments

We investigate whether material variability is critical or not. First, it is supposed to be negligible. Before studying, it should be checked that standard calibration techniques do not present specific difficulties.

[1] For other material configurations, the damage threshold appears as a kink on the strain-stress curves and is thus determined more easily (see SiC-SiC Ceramic Matrix Composite materials in [Zawada et al., 2017]). For the model, it corresponds to situations where the damage evolution exponent p is smaller or equal to 1 and produces a non-smooth transition between the 2 regimes that may create issues in Finite Element Analysis.

The ODM model is calibrated with two repetitions of tensile test, experiments 9 and 10 depicted in Figure 1.3. The noise is supposed to be Gaussian distributed and homoscedastic, so $\boldsymbol{\Psi} = [\boldsymbol{\theta}, \omega]$, giving the following likelihood function:

$$\mathcal{L}(\boldsymbol{\Psi}) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{\Omega}|}} \exp\left(-\frac{1}{2}\left(\mathbf{y} - \mathcal{F}(\mathbf{t}, \boldsymbol{\theta})\right)^\top \boldsymbol{\Omega}^{-1} \left(\mathbf{y} - \mathcal{F}(\mathbf{t}, \boldsymbol{\theta})\right)\right) , \tag{1.25}$$

with $\boldsymbol{\Omega} = \omega^2 \mathbf{I}_N$ (the index i is dropped as both repetitions are considered independently). The minimization of the opposite log-likelihood is performed with the CMA-ES algorithm implemented in [Hansen et al., 2019]. The recommended population size for the CMA-ES algorithm is $\lceil 4 + 3\ln(d+1) \rceil$ [Hansen et al., 2003], which here gives 9 when rounded. Given the model is fast to estimate ($\approx 10^{-4}$ second for one model evaluation), the maximum number of iterations is set to 2000. No constraint is considered except the bounds on the model parameters given in Table 1.2. These bounds are determined from typical values of the model parameters for this kind of materials. The initialization is taken at the center of the hypercube defined by the bounds. The optimization variables are all normalized between 0 and 1 for the minimization of the likelihood given the differences in their scales.

Table 1.2: Bounds used for the model parameters.

	E_0 [GPa]	y_{0s} [kPa]	y_{cs} [MPa]	d_c [-]	ω [MPa]
Upper Bounds	220	5.00	10.0	7.00	4.47
Lower Bounds	120	5.00×10^{-3}	0.1	0.1	0.317

The calibration is achieved for all specimens but for the sake of clarity, we only display in Figure 1.5 the results for 2 experiments (the ninth (later labeled test 1) on the left and the tenth (later labeled test 2) on the right). The numerical values of the calibrated parameters are given in Table 1.3.

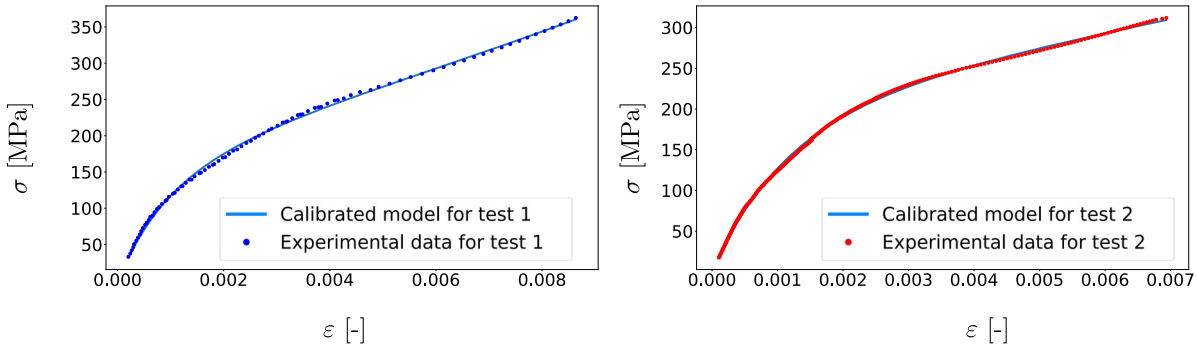


Figure 1.5: Calibration of two tensile tests.

Table 1.3: Numerical values of the calibrated parameters.

	E_0 [GPa]	y_{0s} [kPa]	y_{cs} [MPa]	d_c [-]	ω [MPa]
Test 1	177	5.00×10^{-3} *	2.83	3.75	3.13
Test 2	166	5.00×10^{-3} *	7.49	5.82	2.25

In Table 1.3, the estimated threshold for both experiments hit the lower bound, as signaled by the symbol $*$. In fact, the calibration would like to set it to values that are not physically relevant. It illustrates that this parameter is complicated to estimate properly on the global tensile test response curve, without the help of acoustic data. We can also notice that the

calibrated standard deviation of the noise is below 3.5 MPa, which can be considered as low, given that the stress reaches values over 250 MPa (about 1% of variation around the calibrated curve). To study the convergence of the minimization, various initialization points are tested, namely $\gamma \times \mathbf{1}_{\mathbb{R}^5}$ with $\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. Then, the coefficient of variation, $\mathrm{COV}(X)$, defined for any non-zero quantity X as $\mathrm{COV}(X) = \frac{\sqrt{\widehat{V}(X)}}{\widehat{M}(X)}$ with $\widehat{M}(X)$ the empirical mean of X and $\widehat{V}(X)$ its empirical variance is estimated from the different calibrations and displayed in Table 1.4.

Table 1.4: Coefficient of variations in % of the calibrated parameters for different initializations.

	$\mathrm{COV}(E_0)$ [%]	$\mathrm{COV}(y_{0s})$ [%]	$\mathrm{COV}(y_{cs})$ [%]	$\mathrm{COV}(d_c)$ [%]	$\mathrm{COV}(\omega)$ [%]
Test 1	2.00×10^{-5}	2.97×10^{-7}	4.48×10^{-5}	1.61×10^{-5}	2.02×10^{-4}
Test 2	8.21×10^{-6}	5.96×10^{-8}	5.13×10^{-5}	1.73×10^{-5}	1.20×10^{-4}

Table 1.4 shows that whichever initialization point is chosen, the same parameters are estimated, leading to think that set of parameters calibrated for both test provided in Table 1.3 may be the global minima of the opposite log-likelihood. Note that for the damage threshold, it only indicates that the lower bound is always hit. To study the convergence of the algorithm, we can investigate the values of the model parameters selected by the algorithm. They are depicted in Figure 1.6 for both tests.

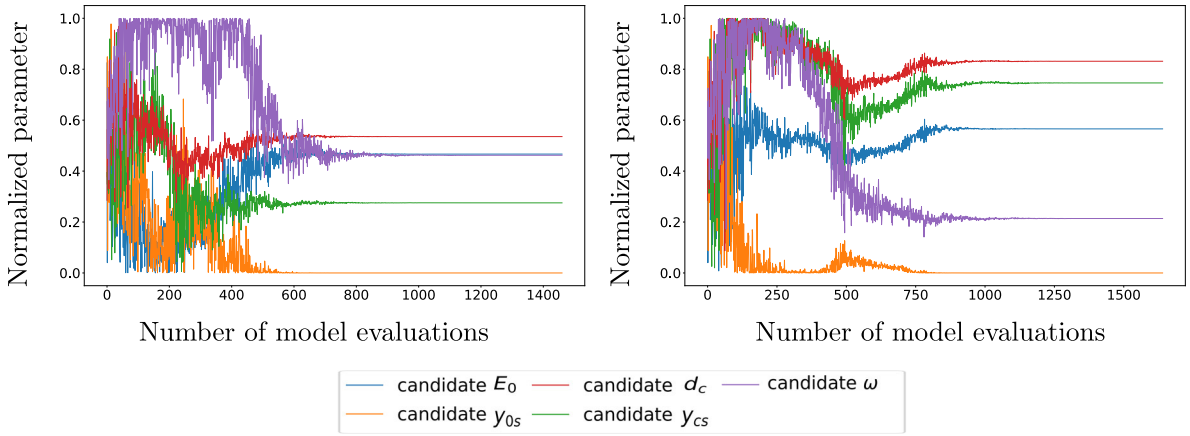


Figure 1.6: Normalized parameters through the optimization for both tests.

It shows that about 1000 iterations are enough to provide a first estimation of the model parameters. Nevertheless, it seems that y_{cs}, d_c and ω take a little longer to converge than E_0, which can be understood as this parameter is the most influential from both a physical and numerical point of view. To finish the study of this first calibration results, we can study the estimated raw residuals for the experiments processed with the calibration which stand for $\widehat{\mathbf{r}}_i = \mathbf{y}_i - \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)$, $i \in \{8, 9\}$. Remember that are supposed to be i.i.d, and drawn from a Gaussian distribution with zero mean and variance ω^2. Plotting the residuals against the fitted values shows gives the picture presented in Figure 1.7.

Figure 1.7 shows that the estimated residuals exhibit a structure. This can be related to presence of model inadequacy. However, for both tests, the magnitude of the residuals remain small with respect to the output value: they represent at most 7% of the model output for test 1 and 5.5% for test 2 and thus do not prevent a proper calibration of the strain-stress curves. This therefore justifies to a certain extent the fact that model bias is neglected. Indeed, this represents a significant increase of the method complexity that should be confronted to low expected gains. The oscillations around the global trend shows the presence of experimental

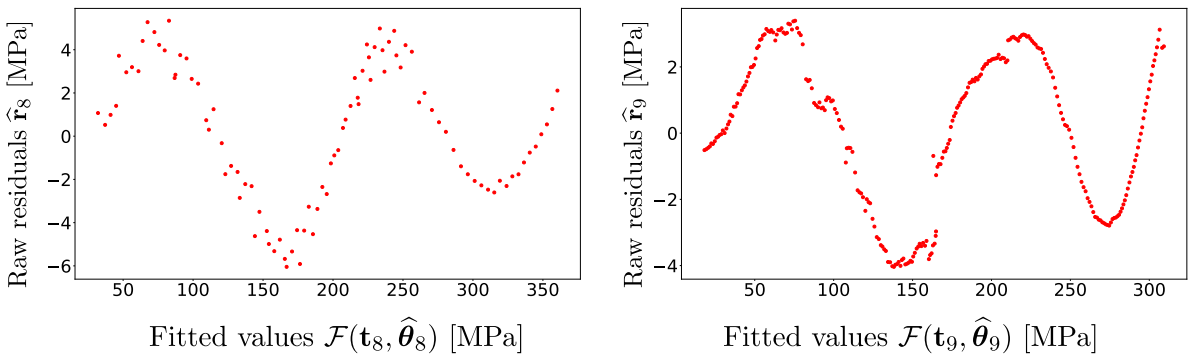


Figure 1.7: Empirical residuals vs. fitted values for Test 1 (on the left) and Test 2 (on the right).

noise. The amplitude of these oscillations seems to be homogeneous across the experiment. Thus, deprived from model inadequacy, the assumption of homoscedastic variance seems to be reasonable. To confirm the non-Gaussian nature of the residuals, one can take a look to the Figure 1.8 that displays the QQ-plot of the residuals for both tests. It shows again that the hypothesis of Gaussian distributed residuals is not compatible with the empirical residuals. However, as the absence of the bias term does not seem to prevent a proper calibration of the experimental data, it will be kept in the following for the sake of simplicity. Note that such trend is found with the other experiments that are not depicted here for the sake of clarity.

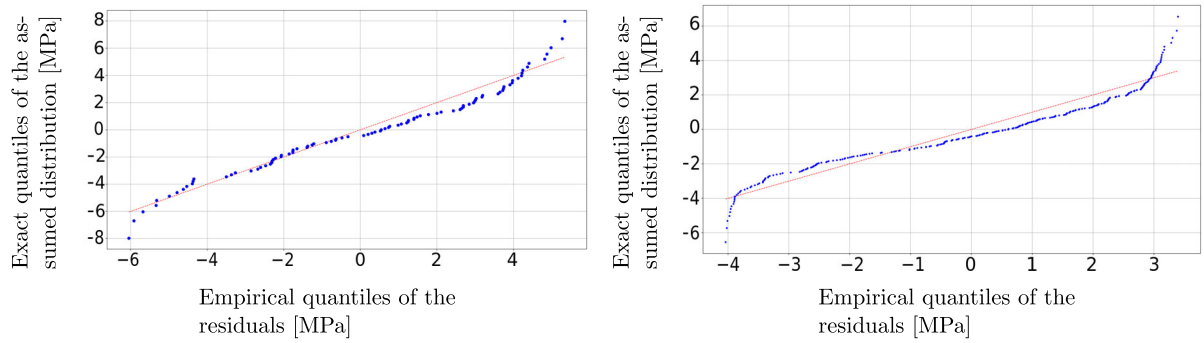


Figure 1.8: QQ-plot of the residuals for test 1 (on the left) and test 2 (on the right).

1.7.3 Highlighting material variability

We investigate whether material variability is critical or not. First, it is supposed to be negligible. Thus, the model parameters ought to be the same (or comparable) between the different experiments, up to the impact of the experimental noise, and in model space, the curves characterizing the impact of uncertainties of two different repetitions should match. Indeed, if we limit ourselves to the deterministic estimates reported in Table 1.3, the model parameters do not seem to be in agreement with each other (except for the damage threshold but this is more a numerical than a physical feature).

Because of model bias and, to a lesser extent, measurements errors, the exact set of model parameters cannot be estimated and their estimates $\widehat{\boldsymbol{\Psi}}$ is flawed with uncertainties. Provided that the residuals are not Gaussian distributed, the asymptotic theory does not apply and the only relevant method to estimate the uncertainties on the model parameters is non-parametric bootstrap. The separation between the two regimes is determined as the moment at which the linear model is no longer accurate to mimic the data as proposed in [Grail, 2013]. The same fitting criterion, the same optimization settings (*i.e.* fitting criterion, bounds, initialization, optimization algorithm, stopping criterion, *etc.*) are considered. The optimal residuals estimated with the first calibration $\widehat{\mathbf{r}}$ for both repetitions are then resampled as described in Section 1.4.2 to form the bootstrapped data that is used for model calibration. Note that residuals from the linear and non-linear regimes are resampled separately. Four thousands bootstrap samples are generated for each experiment.

Uncertainty quantification can also be achieved with Bayesian techniques. We seek to compare the results from both techniques to analyze the consequences of both methodologies. With the likelihood defined in Eq. (1.25) and a uniform prior defined over the bounds detailed in Table 1.2, the posterior density defined in Eq. (1.20) through the Bayes theorem can be sampled with the MH algorithm described in 1. This sampling is carried out with the implementation available in the Python module emcee [Foreman-Mackey et al., 2013] that proposes a variant of the MH algorithm that demonstrates better convergence properties [Foreman-Mackey et al., 2013]. The algorithm runs several chains in parallel. For computational costs reasons, the number of parallel chains is set to 4. The starting points of the chains correspond to $\widehat{\boldsymbol{\Psi}} \times (\mathbf{1}_{\mathbb{R}^5} + \boldsymbol{\delta})$, with $\boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}_{\mathbb{R}^5}, 10^{-4}\mathbf{I}_5)$ and $\widehat{\boldsymbol{\Psi}}$ the MLE estimate. Each chain generates 4000 samples.

Before comparing the different methods, it is necessary to assess the convergence of MCMC algorithm. Though it is complicated to exactly diagnose this convergence, it is possible to check the plot describing the values of the parameters visited by the sampling algorithm, as depicted in Figure 1.9.

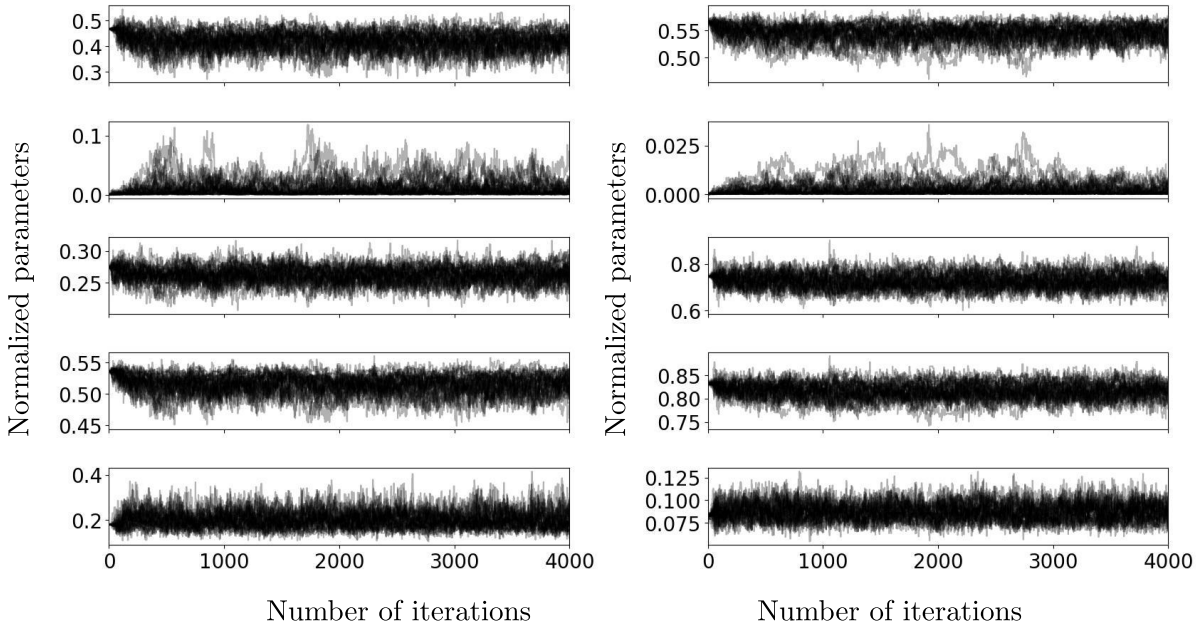


Figure 1.9: Markov Chains for the MCMC sampling for normalized parameters $\boldsymbol{\Psi}$. The first row displays the chains for the normalized modulus, the second for the normalized damage threshold, the third to the normalized damage evolution celerity, the fourth to the normalized damage saturation and the last to the normalized standard deviation. The columns correspond to the two tests studied.

In Figure 1.9, it seems that after 500 iterations, the chains have reached a stationary behavior and explored most of the support of the posterior distribution. Note that, as for standard calibration, the samples of damage threshold hit the lower bound. This still comes from the low sensitivity of the model output with respect to this model parameter. Furthermore, the estimated threshold are not physically consistent because they hint for the absence of linear behavior contrary to what the data suggest. Before exploiting the samples, it is necessary to burn the first samples and thin the remaining ones. A conservative choice is to burn the first 1000 samples. To restore independence between the rest, one every 30 samples available is kept (thinning). In total, this makes 4000 samples available for the posterior distribution. To compare the bootstrap and Bayesian methods, it is possible to plot the empirical distributions of the bootstrapped and MCMC samples, as it is done in Figure 1.10 for test 2.

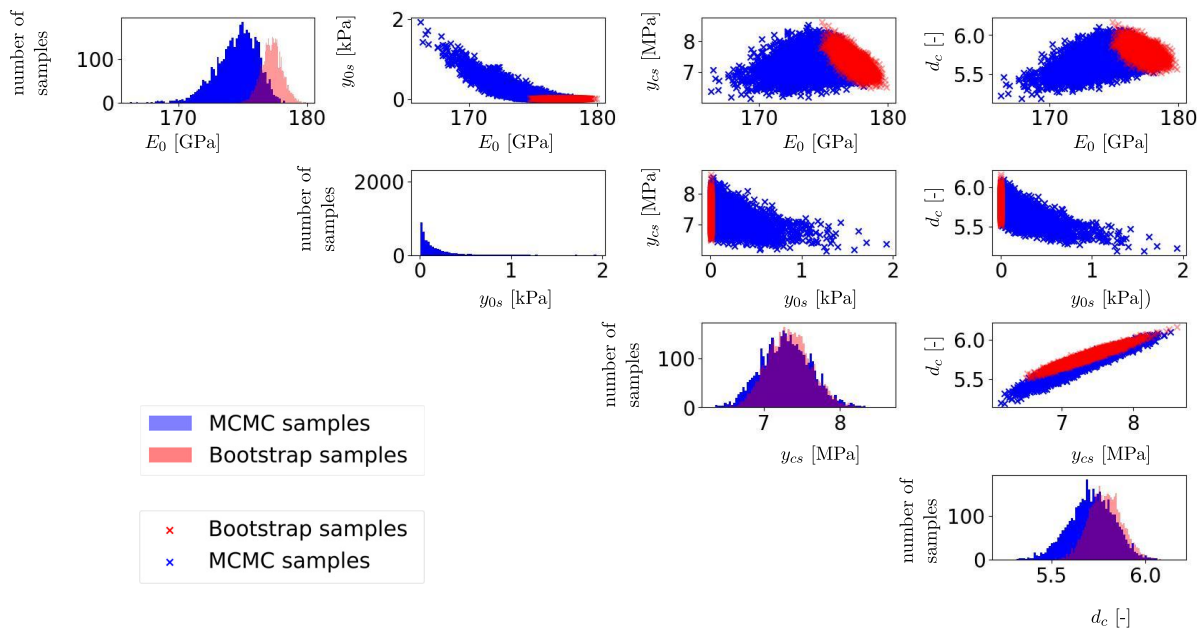


Figure 1.10: Empirical distributions for test 2 with bootstrap and MCMC techniques.

Let us first concentrate on the joint distribution of y_{cs} and d_c, the damage evolution celerity and the damage saturation, respectively. Both techniques provide similar results in terms of marginals and dependence structure. Yet, they provide different results for the other marginals and dependence structure. Note that the bootstrap samples for the damage threshold are all stuck to the lower bound. It is difficult to interpret the results further for the damage threshold given the small sensitivity of the model output with respect to this parameter such as the false and skewed distribution of the threshold yielded by this pair of model and experiment. Notice also that MCMC allows a more complete coverage of the support of the empirical distribution of the model parameters than bootstrap. It seems that with "only" 4000 samples, the bootstrapped data did not exhibit sufficient variety to cover the whole support of the empirical distribution of $\mathbf{\Psi}$. Another difference lies in the fact that bootstrap samples are determined after likelihood minimization, meaning they represent the "best" achievable fit, when Bayesian calibration enables to select points that are not necessary the most relevant with respect to the fitting criterion. This enables a more complete coverage of the support of the distribution of $\mathbf{\Psi}$. In Figure 1.10, we can see that dependencies between the model parameters arise (in the extra-diagonal pictures). This dependence structure accounts for the couplings that appear between the model parameters because of the model formulation. The coupling between d_c and y_{cs} for instance can be understood as follows: if damage develops faster (with speed characterized by the inverse of y_{cs}), then its saturation monitored by d_c is reached faster to keep the same level of error. Another interpretation is: given a fixed E_0, an increase of y_{cs} means a slower damage. Then, the effective modulus is higher than observations suggest. The only way to compensate this change resides in the increase of the damage saturation, as illustrated in the Figure 1.10.

The samples from both MCMC and bootstrap can be propagated through the model. Provided that MCMC samples have a larger support than the bootstrap samples, we choose to display their propagation in Figure 1.11.

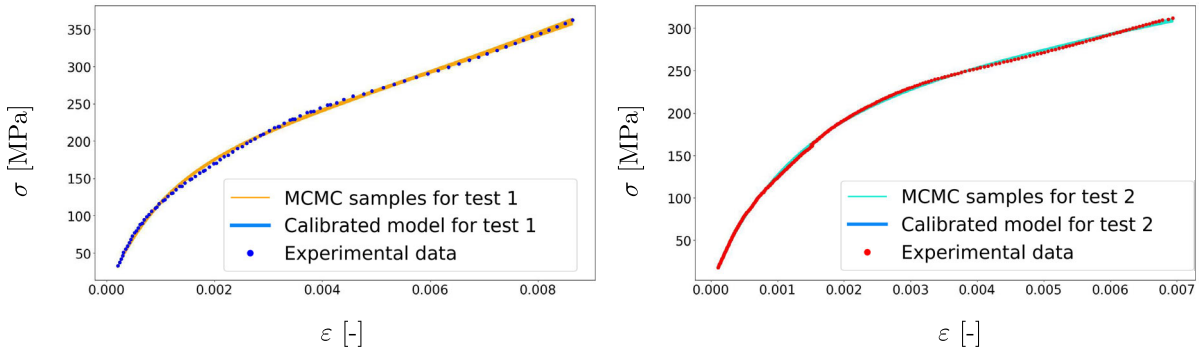


Figure 1.11: Uncertainty quantification for two repetitions of tensile tests.

Propagated into model space, MCMC samples yield the bundles of curves depicted in Figure 1.11. It is interesting to notice that the uncertainty that comes from both model inadequacy and measurements errors is, in model space, very small. To illustrate this fact, we can estimate after the uncertainty propagation an upper (at 97.5%) and a lower (at 2.5%) quantiles from the curves at each observation point (which gives an 95% confidence interval for the output stress). The difference between the upper and lower quantiles reaches at most 5 MPa for the first test and 2.3 MPa for test 2, compared to the levels of stress that go over 300 MPa.

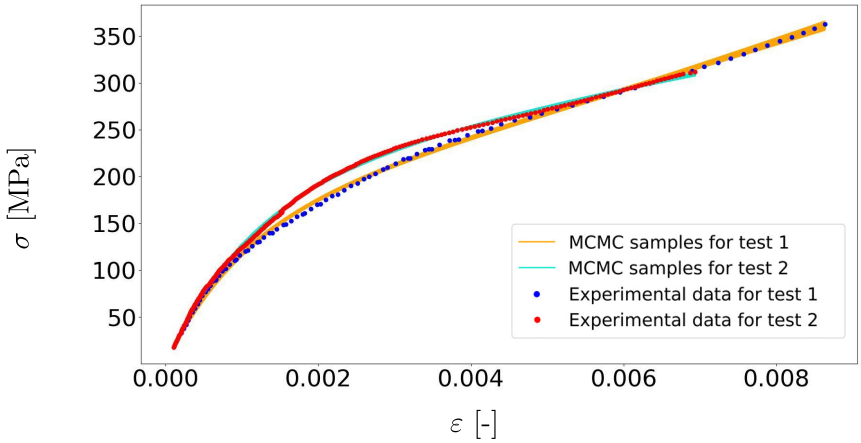


Figure 1.12: Uncertainty quantification for two repetitions of tensile tests.

Because of this small variation around the mean curve, we can see the identification of the two tests are not compatible with each other. To be more precise, the bundle of model outputs for the MCMC samples from test 1 does not match with the bundle of outputs from test 2. In fact, Figure 1.12 shows that neither experimental noise nor model bias can explain the gap between the two different tests. This can also be depicted in parameter space by comparing the empirical distributions of the MCMC samples illustrated in the parameters space.
The incompatibility between the parameters charactering each specimen in model space in Figure 1.12 is depicted in the parameter space in Figure 1.13. Globally, it can be noticed that the support of the empirical distributions of the MCMC samples for both tests are disjointed for all parameters, even if there is a small intersection fot the Young's modulus, that still is located for the upper and lower bound and is thus statistically not representative. The distributions of the damage threshold exhibit the same behavior as earlier. This phenomena can be numerically quantified. Given the MCMC samples $\{\boldsymbol{\theta}_{\ell,2}^{\mathrm{MCMC}}\}_{\ell\in[\![1,4000]\!]}$ from test 2, thanks to kernel smoothing [Wand and Jones, 1994], their empirical distribution can be estimated. It is possible to compute the log-PDF $\ln(\widehat{f}(\cdot))$ of this distribution for the set of model parameters calibrated from the first and second experiment, $\ln(\widehat{f}(\widehat{\boldsymbol{\theta}}_1))$ and $\ln(\widehat{f}(\widehat{\boldsymbol{\theta}}_2))$, respectively equal to -10^{308} (a numerical threshold for $-\infty$) and 2.33. It shows that it is unlikely for $\widehat{\boldsymbol{\theta}}_1$ to either explain the data of experiment 2 or belong to the support of the empirical distribution of the MCMC samples. The same conclusion can be drawn if the MCMC samples are replaced by the bootstrap samples. To

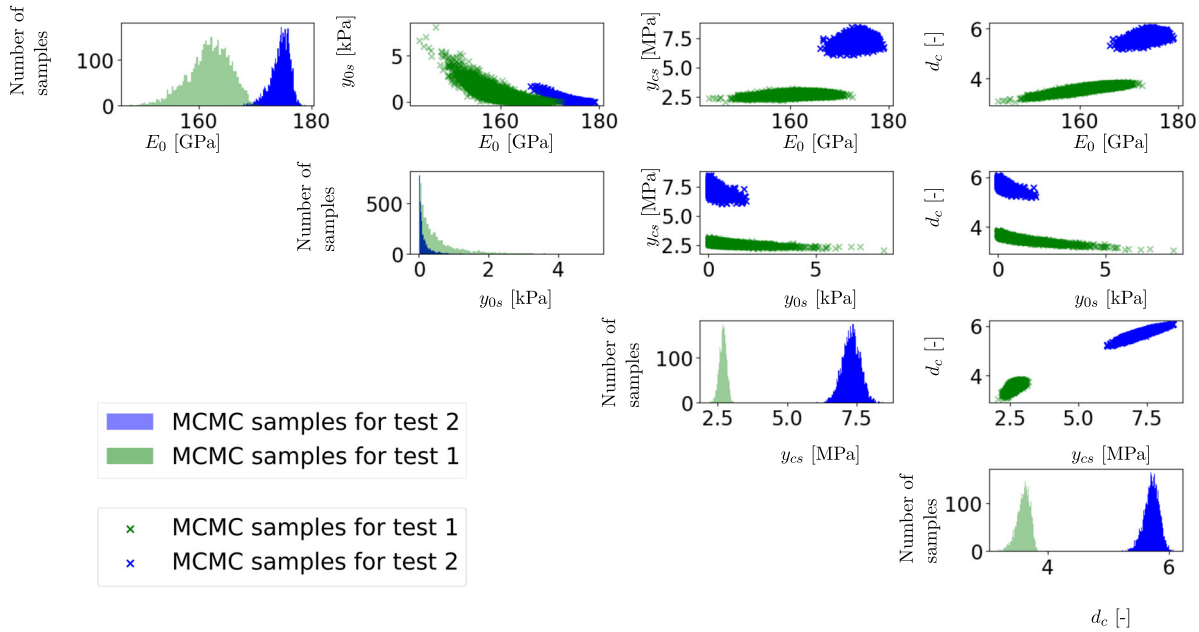


Figure 1.13: Empirical distributions of the MCMC samples for test 1 (in green) and test 2 (in blue).

summarize, there are incompatibilities between the two experiments that cannot be explained either by measurements errors or model inadequacy provided that MCMC samples are supposed to account for such uncertainties. One possible explanation to this discrepancy is the presence of significant material variability. Note that also, given that the impact of both measurements errors and model bias is much smaller than the intrinsic variability, for the sake of simplicity, the residual model considered so far can be retained. This justifies subsequently to keep the same assumptions on the residual model, even if they do not fully agree with the observations.

1.7.4 Estimating material variability with multi-objective techniques

Because of material variability, the bootstrap samples for one experiment are not consistent with the observations of other repetitions. To get model parameters relevant to both repetitions at the same time, it is possible to use multi-objective calibration, with each objective being associated to a repetition [Madsen, 2003]. The provided trade-offs partly come from material variability. Thus, the Pareto front characterizes the impact of material variability among other sources of uncertainty, and material variability can be estimated from the corresponding model parameters.

The estimation of material variability in a multi-objective fashion is carried out in the following. For two repeated experiments, the objectives considered here are:

$$\mathcal{C}(\boldsymbol{\Psi}) \coloneqq [-\ln(\mathcal{L}_1(\boldsymbol{\Psi})), -\ln(\mathcal{L}_2(\boldsymbol{\Psi}))] \ , \tag{1.26}$$

with $\mathcal{L}_i$ the likelihood associated to the i^{th} experiment. The minimization of the objective defined in Eq. (1.26) is performed with the NSGAII [Coello, 2000] algorithm implemented in the Platypus package [Hadka, 2015]. The research space is limited to the bounds defined in Table 1.2. No constraints are taken here into account. The population size is set to 2000 to generate significant variety of the candidate samples. Other settings are kept to their default values. The stopping criterion is the maximum number of iterations. To test the convergence of the optimization, we try several numbers of iterations of the NSGAII algorithm, namely $10^4, 5{\times}10^4, 10^5$ and $2{\times}10^5$ (that can be considered as the upper limit of acceptable computational costs with the available facilities). Note that the minimization is not performed over the variance of the noise. From Figure 1.14, it appears that with 10^5 iterations, the estimation of the Pareto

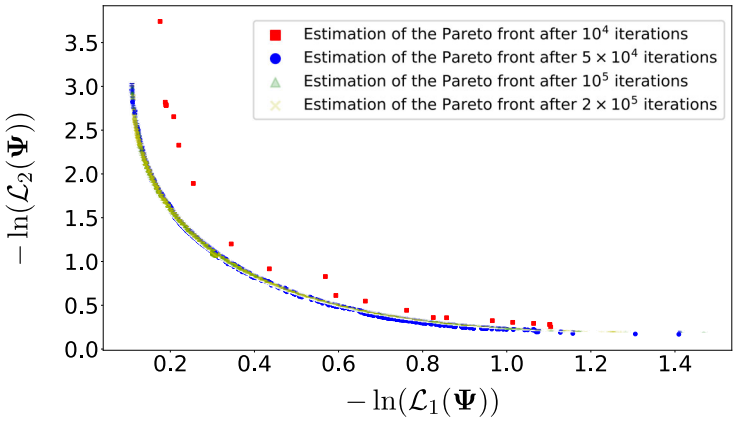


Figure 1.14: Different estimations of the Pareto front with $10^4, 5\times10^4, 10^5$ and 2×10^5 iterations of the NSGAII algorithm.

front has converged and adding iterations does not provide additional information. Vice versa, 10^4 iterations are not sufficient to estimate the Pareto front in terms of coverage, uniformity and number of non-dominated candidate samples. Increasing the budget to 5×10^4 almost allows to get to the Pareto front. Yet, its description lacks continuity that is achieved with both 10^5 and 2×10^5 iterations. Note that for $5\times10^4, 10^5$ and 2×10^5 iterations, the number of non-dominated samples of the final population is 2000, the same as the initial population size.

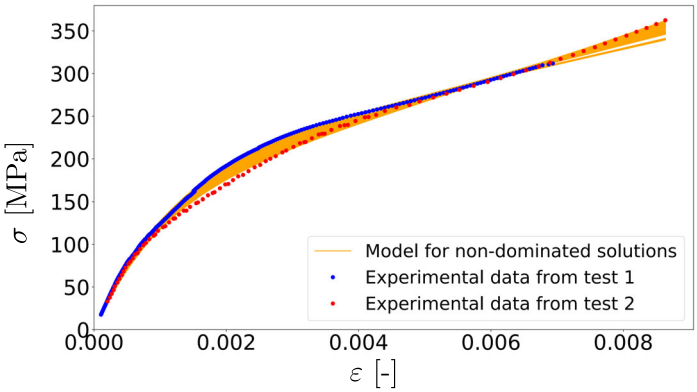


Figure 1.15: Models for non-dominated solutions of the multi-objective optimization.

The results from the multi-objective optimization are depicted in Figure 1.14. First, it is possible to notice the convex shape and the significant interval of variation of the Pareto front. This indicates the presence of trade-offs between the two objectives and underlines the presence of material variability. Furthermore, it can be checked that model predictions for the provided non-dominated points are consistent with both experiments as illustrated in Figure 1.15. Indeed, the discrepancy between the two experiments is now filled with model predictions. The objective to get trade-offs can be considered as fullfilled. The repartition of the candidate samples is illustrated in Figure 1.16. As with previous techniques, it can be noticed that the damage threshold hits the lower bound. The coupling between y_c and d_c appears again. Yet, the difficult point is to know whether these candidate samples can be interpreted as being drawn from a probability distribution that has a physical interpretation. Such interpretation is not reached because that the relation of Pareto-dominance is deterministic and does not include any random aspect. In addition, even if this interpretation was available, provided that the trade-offs that appear in the Pareto front account for all types of uncertainties, it would be difficult to extract specifically the impact of material variability. For instance, it is difficult to know what is due to experimental noise, model bias or material variability.

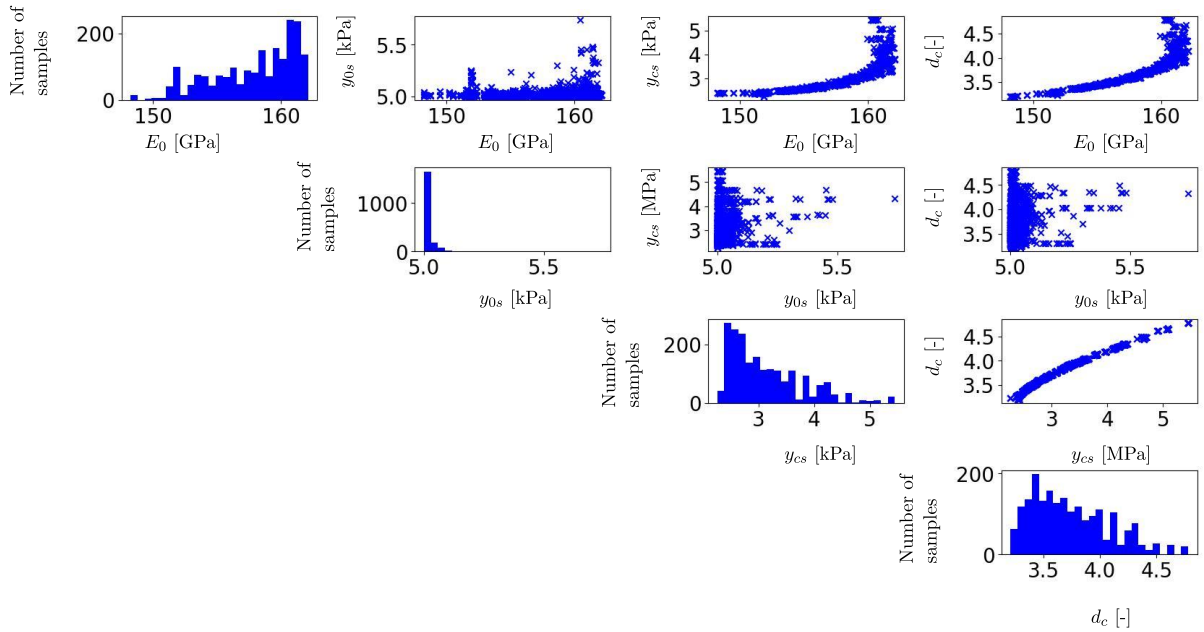


Figure 1.16: Repartition of the candidate samples in the Pareto set of the bi-objective optimization of two strain-stress experiments.

1.8 Partial Conclusions

Considering the large scope of its applications, uncertainty quantification for model calibration is an active field research. The first part of this chapter consists in a literature review of the usual calibration methods and was limited to the standard ones. The statistical modeling presented here is sufficiently general to apply in different situations. In particular, it provides a modeling and the associated methods to achieve the calibration of repeated experiments. It can be noticed that all methods allow to propagate the uncertainties to the estimated parameters therefore yielding associated dispersions.
Nevertheless, these methods do not provide satisfactory estimates of material variability. It is a key issue as it is known to be critical for composite materials. This has been illustrated in the second part of this chapter on repetitions of tensile tests on a woven ceramic matrix composite. An attempt to characterize material variability through multi-objective optimization was carried out but suffers from several limitations. For instance, within the Pareto front, the impact of all uncertainties are mixed together making it almost intractable to extract specifically the effect of material variability.
In fact, the main problem of the presented approach is the assumption of a unique set of model parameters for all the repetitions. Consequently, we choose to focus our research on another framework, the population approach [Demidenko, 2013, Lavielle, 2014] described further in the next chapter. This framework comes from the pharmacokinetics field and proposes to consider different model parameters for each repetition. Variability is modeled by a probability distribution whose outcomes are the set of parameters, both of which will be estimated.

CHAPTER 2

STATE OF THE ART OF THE POPULATION APPROACHES

Contents

Chapter goals:

- Present the population approach
- Highlight of the differences with the individual approach described in the first chapter (in particular on the statistical modeling)
- Introduce and compare two usual calibration methods: hierarchical Bayesian Models and mixed-effects models
- Describe the methods to estimate the population parameters based on the maximization of the likelihood of mixed-effects
- Report the proper ways to maximize the likelihood of mixed-effects
- Explain the possible analyses of a mixed-effects calibration result

2.1 Introduction

The determination of material variability has already been studied in the context of mechanical problems. The word "variability" refers here to the dispersion of the material properties that can be observed throughout repetitions of specimens. Several works in the field of material science tried to characterize the variability of mechanical properties ([Currey et al., 2007] for human bone, [Niyigena et al., 2016] for hemp concrete and [Bright et al., 2011a] for high strength low alloy steel). Generally, there are two different trends. First, from measurements or databases available in the literature [Lekou and Philippidis, 2008, Maljaars et al., 2022], the variability modeled with a distribution can be inferred. This applies to properties that are directly observed from experiments (*e.g.* ultimate strengths, densities) or determined from the observations according to standards (*e.g.* Young's modulus) [An et al., 2021, Currey et al., 2007, Murugan et al., 2008]. A second possibility is to exploit the multi-scale behavior of composite materials to determine the meso-scale properties from either micro or macro scale properties [Sriramula and Chryssanthopoulos, 2009, Toft et al., 2013, Vinot et al., 2023]. However, in many material applications, most of the model parameters can be neither estimated directly from the data nor computed using multi-scale models. In fact, the only option left is to calibrate a behavior law which takes as inputs these parameters to get their estimate.
Chapter 1 illustrated the inadequacy of the usual methods [Gallagher and Doherty, 2007, Kennedy and O'Hagan, 2001] to estimate the impact of material variability on the parameters of a simplified ODM-type material model (that are not observed directly). In the previous chapter, we have made a first attempt using multi-objective calibration [Efstratiadis and Koutsoyiannis, 2010, Vrugt et al., 2003] to characterize material variability from repetitions of tests. Insofar, the main issue encountered with multi-objective calibration is that within the Pareto front, all types of uncertainties, including material variability, are combined together, making it difficult to extract the impact of a specific one. To estimate material variability, a broader framework has been proposed in [Ballesteros et al., 2014, Song et al., 2020]: the population approach [Beal and Sheiner, 1982, Fisher, 1919]. This approach originates in biological studies and aims at describing the intrinsic variability of subjects submitted to the same test (*e.g.*, the injection of a drug in pharmacometrics [Ette and Williams, 2004]). This kind of approaches has also been applied in agriculture [Hall and Clutter, 2004] or epidemiology [Wang et al., 2020] among others. Contrary to the methods presented in the first chapter that could be labeled as "individual" approaches [Lavielle, 2014], in the population approaches, the assumption of a single set of values of model parameters no longer holds and the different subjects involved in

each repetition are described by different values of the model parameters. The subjects stand here for the repetition of specimens on which are carried out the experiments. Considering this fact, population approaches intend to provide estimates both of the population overall behavior and of the inter-individual variability. A first illustration of such a characterization can be found in [Jacquet et al., 2017]. Different population approaches exist in the literature, for example Hierarchical Bayesian Modelings or mixed-effects models. As explained in Section 2.2, these methods differ in the way to estimate the variability (using MCMC for HBM or by likelihood maximization for mixed-effects). In fact, population approaches should be viewed as a generalization of the individual approaches presented in the first chapter. Indeed, if the inter-individual variability is negligible, then, the sets of values of the model parameters describing each specimen become identical, which is the key statement of the individual approach.
The purpose of this chapter is to first make a panorama on population approaches (Section 2.2) then to focus on mixed-effects models as more appropriate for our study as discussed in Section 2.2.4. For that purpose, Section 2.3 presents the cornerstone of the mixed-effects approach that is the formulation of the likelihood maximization problem to enable the calibration the model. More precisely, Section 2.3.1 presents its derivation and the different methods used for its estimation and maximization are discussed in Sections 2.3.2 and 2.3.3. Still, the likelihood maximization can also be carried out by developing sequential procedures that simplify the optimization process. These alternate formulations are presented in Section 2.3.4. Finally, the analysis of the results from mixed-effects calibration of phenomenological model is discussed, including the presentation of appropriate statistical indicators in Section 2.4. The methods and notions presented in this literature review are analyzed through the scope of our application context.

2.2 Overview of the population approaches

2.2.1 Main hypotheses

In the previous chapter, the study of repetitions of tensile tests performed on a woven ceramic composite showed that material variability may be significant. Because of the assumption of a unique set of model parameters for all experiments (as illustrated on the left of Figure 2.1), material variability is mixed together with measurements errors and model inadequacy. A broader framework in which the intrinsic variability of the material properties is explicitly modeled (that is to say where the different specimens are characterized by different set of model parameters values) has been proposed in [Congdon, 2007, Lavielle, 2014].
Changing this key assumption leads us to update that statistical formalization of the calibration problem. The population approaches aim to characterize the effects of individual variability over a population of specimens, also referred to as individuals [Demidenko, 2013, Lavielle, 2014]. They provide an estimation of the intrinsic variability of the properties alongside an approximation of the parameters describing each specimen. They assume that because of this variability, the different samples (n in total) are characterized by distinct sets of model parameters values denoted $\boldsymbol{\theta}_i, i \in [\![1, n]\!]$. These sets of model parameters are different from experiment to experiment (*i.e.* $i \neq j \Rightarrow \boldsymbol{\theta}_i \neq \boldsymbol{\theta}_j$) and are called the individual parameters. No spatial and temporal variation of the parameters throughout the experiment is considered here again for the sake of simplicity. In a probabilistic framework, this variability can be modeled by a probability distribution $f_{\boldsymbol{\Theta}}$ whose outcomes are the individual parameters: $\boldsymbol{\theta}_i \sim f_{\boldsymbol{\Theta}}$, with $\boldsymbol{\Theta}$ the random variable associated to the $\boldsymbol{\theta}_i$s. If $f_{\boldsymbol{\Theta}}$ is parametric (a joint Gaussian probability distribution for instance), let $\boldsymbol{\Pi}$ be its parameters (also called the population parameters) and $f_{\boldsymbol{\Theta}} = f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$. The objective of the method is to estimate both the model parameters distribution $\boldsymbol{\Pi}$ and the individual parameters $\boldsymbol{\theta}_i$. If $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ is a parametric distribution, identifying $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ is tantamount to determining $\boldsymbol{\Pi}$. Figure 2.1 sums up the differences between the assumptions in the individual (left of the figure) and population (right of the figure) approaches.

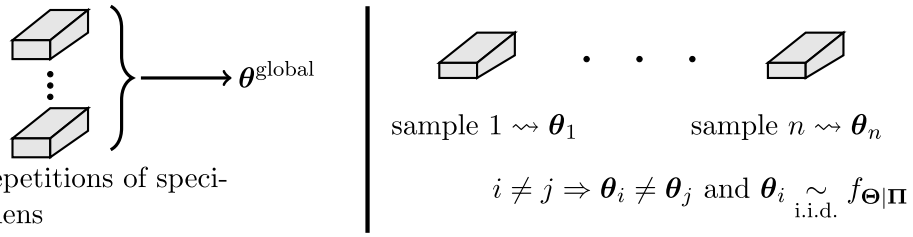


Figure 2.1: Two hypotheses for the calibration of repeated experiments. On the left, all experiments are described by a single set of the model parameters. On the right, each specimen is characterized by distinct values of the model parameters.

The population approaches have been applied successfully in many different fields where the consideration of the intrinsic variability between repetitions of specimens is key feature to provide reliable characterization of the individuals properties. Such approaches have been introduced in biological studies [Fisher, 1919] in which it is important to describe the specificity of each individual. Among those can be found pharmacometrics [Drikvandi, 2017, Ette and Williams, 2004], economics [Nelson and Siegel, 1987, Svensson, 1994], agriculture [Baey et al., 2016, Hall and Clutter, 2004], oil & gas industry [Harris and Lee, 2014, Lewis and Beal, 1918], epidemiology [Wang et al., 2020] and applications in structural mechanics [Ballesteros et al., 2014, Song et al., 2020]. In pharmacometrics, the use of such methods is linked to the inherent variability of the living subjects on which the experiments are carried out (*e.g.*, drug injection) because of biological reasons (genetics, living conditions, *etc.*) In structural mechanics, accounting for the variability of civil engineering structures is major issue in order to ensure a proper sizing. This variability partly comes from the behavior of the base materials (concrete, soils, *etc.*) which can be modified by temperature and humidity conditions [Song et al., 2020], leading to a significant amount of uncertainty in the behavior of these materials.

Having introduced the population approach, the objective of next section is to describe the corresponding statistical modeling.

2.2.2 Statistical formalization of the population approach

Formally, the probability space in which $\boldsymbol{\Theta}$ evolves, the random vector whose realizations are the individual parameters, is denoted $(\Gamma_{\boldsymbol{\theta}}, \mathcal{T}_{\boldsymbol{\theta}}, \mathbb{P}_{\boldsymbol{\theta}})$. Here, $\mathbb{P}_{\boldsymbol{\theta}}$ refers to the probability measure that models material variability (and not the impact of measurements errors as in the previous chapter). The population approaches framework [Demidenko, 2013, Lavielle, 2014] assumes that there exists a probability distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ whose outcomes are the individual parameters:

$$\forall i \in [\![1, n]\!],\ \boldsymbol{\theta}_i \underset{\text{i.i.d.}}{\sim} f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}\ , \tag{2.1}$$

with i.i.d. standing for independently identically distributed and n the number of available specimens. The individual parameters of dimension d belong to $\Gamma_{\boldsymbol{\theta}}$, a subset of $\mathbb{R}^d$ that represents mathematical and physical constraints on the model parameters. Eq. (2.1) describes the population level [Davidian and Giltinan, 1995]. We refer to $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ as the model parameters probability density function (PDF). In a calibration context, both $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ and the $\boldsymbol{\theta}_i$s are unknown and the aim is to determine them. Throughout this manuscript, it is taken as a parametric distribution and let $\boldsymbol{\Pi}$ be its parameters (hence the dependency of $\boldsymbol{\Theta}$ with respect to $\boldsymbol{\Pi}$) with s its dimension. Identifying $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ is equivalent to determining $\boldsymbol{\Pi}$.

$\boldsymbol{\Pi}$ has an important interpretation. Notably, it can describe physical couplings between the material parameters $\boldsymbol{\theta}$. It is important to note that these dependencies do not refer to those induced by the statistical model formulation but by the physics. To illustrate notion of physical dependencies, remember that composite materials are multi-scale materials [Berthelot, 2012].

More specifically, laminated composites match with a stacking sequence of base plies, themselves composed of fibers and matrix, whose behavior is independent as they come from different materials. Consequently, the parameters that describe phenomena mostly driven by the fiber behavior (longitudinal elasticity) and those that describe phenomena mostly driven by the matrix behavior (transverse elasticity) can be considered as physically independent. On the contrary, the parameters describing the phenomena that combine both fiber and matrix behavior (shear properties) should be considered as dependent [Berthelot, 2012]. Note that they do not correspond to spatial couplings following the assumed absence of spatial of the material properties within an individual. Formally, the couplings between the coordinates of random vector $\boldsymbol{\Theta}$ can be modeled in PDF $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ with the help of copulas. Remember that from the Sklar's theorem [Sklar, 1959], any PDF can be (uniquely) decomposed between marginals and copulas as:

$$f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) = c(F_1(\theta_{i1}|\boldsymbol{\Pi}_1), \dots, F_d(\theta_{id}|\boldsymbol{\Pi}_d)|\boldsymbol{\Pi}_{\text{cop}}) \times \prod_{l=1}^{d} f_l(\theta_{il}|\boldsymbol{\Pi}_l) \ , \tag{2.2}$$

with $\boldsymbol{\Pi}_{\text{cop}}$ the copulas parameters, $\boldsymbol{\Pi}_l$, $l = 1, \dots, d$ the parameters of the l^{th} marginal with probability density function (PDF) $f_l(\cdot)$ and cumulative density function (CDF) $F_l(\cdot)$, θ_{il} the l^{th} coordinate of the i^{th} individual parameter $\boldsymbol{\theta}_i$, $(i,l) \in [\![1,n]\!] \times [\![1,d]\!]$ and $c(\cdot)$, the density function of the chosen copula. The marginal and copulas parameters comply with $\boldsymbol{\Pi} \coloneqq [\boldsymbol{\Pi}_1, \dots, \boldsymbol{\Pi}_d, \boldsymbol{\Pi}_{\text{cop}}]$. For example, for a Gaussian random vector, the marginals are described by means and standard deviations contained in $\boldsymbol{\Pi}_1, \dots, \boldsymbol{\Pi}_d$. The dependence structure of $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ is characterized by a Gaussian copula, and $\boldsymbol{\Pi}_{\text{cop}}$ will contain the correlation parameters which describe the modeled interactions. Given $\boldsymbol{\Pi}$ and $\boldsymbol{\theta}_i \sim f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}\ \forall i \in [\![1,n]\!]$, the experimental output $\mathbf{y}_i$ of the i^{th} individual can be written as

$$\forall i \in [\![1,n]\!],\ \mathbf{y}_i \underset{\text{i.i.d.}}{\sim} \mathcal{F}(\mathbf{t}_i, \boldsymbol{\theta}_i) + \mathbf{r}_{\mathbf{t}_i} \ , \tag{2.3}$$

with $\mathcal{F}(\cdot,\cdot)$ the behavior law, $\mathbf{t}_i$ the input data for the i^{th} individual, $\mathbf{r}_{\mathbf{t}_i}$ (simplified into $\mathbf{r}_i$ in the following) the concatenation of all the residuals $r_{t_{ij}}$ (denoted r_{ij} for the sake of clarity) $\mathbf{r}_i = \{r_{ij}\}_{j \in [\![1,N_i]\!]}$ and N_i the number of observation points of the i^{th} specimen. This equation describes the individual level [Davidian and Giltinan, 1995]. As illustrated in Section 2.3 of Chapter 1, the impact of material variability proved to be more significant than both measurements errors and model bias. For the sake of simplicity, we choose to assume that model bias is negligible, so the residuals simplify to the measurements errors, that is to say $\mathbf{r}_i = \boldsymbol{\xi}_i, i \in [\![1,n]\!]$, with $\boldsymbol{\xi}_i = \{\xi_{ij}\}_{j \in [\![1,N_i]\!]}$ the concatenation of all measurements errors for the i^{th} subject. This is consistent with the choices made in the first chapter: neglected bias and Gaussian white noise with homoscedastic unknown variance different for each individual. Without any other hypothesis, the ξ_{ij}s are different for each individual and for each observation. The global population approaches model for the j^{th} output measure of the i^{th} individual y_{ij} reads as:

$$y_{ij} = \mathcal{F}(t_{ij}, \boldsymbol{\theta}_i) + \xi_{ij} \ . \tag{2.4}$$

Another hypothesis has to be made on the error term. To remain consistent with the choices made (Gaussian centered distributed residuals with the same variance for all observations) in the first chapter, a similar assumption is made, which is that for each individual and each measure, the error term is a Gaussian white noise (without correlation):

$$\xi_{ij} \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \omega_{ij}^2) \ , \tag{2.5}$$

with ω_{ij} the standard deviation of the noise of the j^{th} output measure of the i^{th} individual. The assumption of a Gaussian white noise is widely used in practice [Lavielle, 2014], but needs to be validated afterwards (with the QQ-plot of the residuals for instance). Furthermore, the noise is supposed to be homoscedastic for all individuals, *i.e.* $\omega_{ij} = \omega_i\ \forall\ (j,i) \in [\![1,N_i]\!] \times [\![1,n]\!]$ (to

be also validated after calibration). This assumption is interesting because, in such a case, the residual model is characterized with few parameters, which decreases the number of parameters to be estimated.
Finally, the vector of parameters to calibrate, that we name the full population parameters (because they also include the residual models) with p its dimension, is denoted $\boldsymbol{\Psi}$:

$$\boldsymbol{\Psi} \coloneqq \left[\boldsymbol{\Pi},\ \ \omega_1,\ \ \ldots\ \ ,\omega_n\right] = \left[\boldsymbol{\Pi}, \boldsymbol{\omega}\right] .$$

Other assumptions on the noise distribution can be chosen (*e.g.* Weibull distributed noise). In that, case, the ω_i^2s are replaced by the relevant parameters in the definition of $\boldsymbol{\Psi}$.

The population approaches seek to calibrate $\boldsymbol{\Psi}$ and to provide an estimate of the individual parameters $\{\boldsymbol{\theta}_i\}_{i\in[\![1,n]\!]}$ as a by-product. To summarize, the population-based models comply with the following equations:

$$\begin{cases} \forall i \in [\![1,n]\!], \boldsymbol{\theta}_i \underset{\text{i.i.d.}}{\sim} f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}} \ , & \text{(2.6a)} \\ \forall i \in [\![1,n]\!], \mathbf{y}_i \sim \mathcal{F}(\mathbf{t}_i, \boldsymbol{\theta}_i) + \boldsymbol{\xi}_{\mathbf{t}_i} \ , & \text{(2.6b)} \\ \forall i \in [\![1,n]\!], \boldsymbol{\xi}_{\mathbf{t}_i} \underset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_{\mathbb{R}^{N_i}}, \omega_i^2 \mathbf{I}_{N_i}) \ . & \text{(2.6c)} \end{cases}$$

Figure 2.2 illustrates that the different levels of modeling associated with Eqs. (2.6) exhibit a hierarchical structure. Indeed, starting with the population level that describes the intrinsic variability of the mechanical properties characterized by $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$, realizations of individual parameters can be generated according to $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$, see Eq. (2.6a) ($n$ of them here). For each of these individuals, we can continue to the lower, individual, level. Given the distribution of the measurements errors, it is possible to draw profile of noise measurements (Eq. (2.6c)). Combined with the behavior law output estimated with the individual parameters and the associated input data, it becomes possible to model the observed experimental data through Eq. (2.6b).

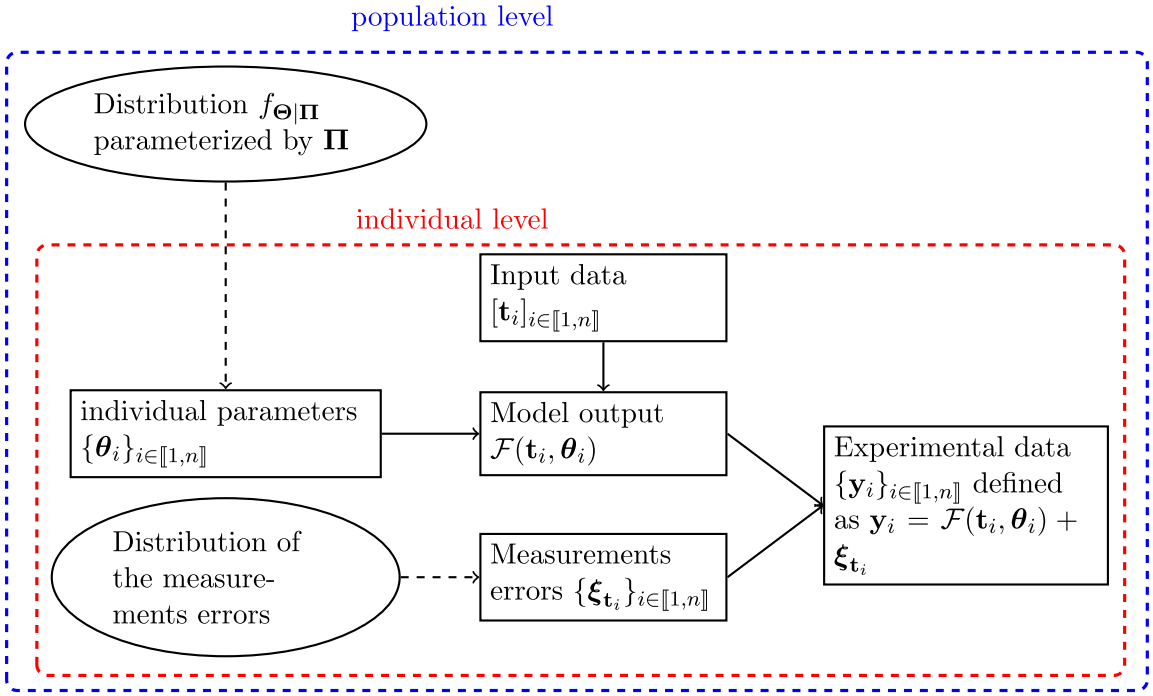


Figure 2.2: Different levels of modeling in the population approach (adapted from [Lavielle, 2014] and [Nagel and Sudret, 2016]). The dashed lines indicate random sampling and the ellipsis the distributions.

To solve the calibration problem, both the population parameters $\boldsymbol{\Pi}$ jointly with the residual models $\boldsymbol{\omega}$ (so $\boldsymbol{\Psi}$) and the individual parameters (the $\boldsymbol{\theta}_i$s) should be estimated. Among the available options, it is possible to exploit the hierarchical structure implied by the population modeling. Thus, the estimation of the population parameters can be carried out in the framework of hierarchical-based techniques, in particular Hierarchical Bayesian Models (HBMs) [Congdon, 2007]. HBMs is a fully Bayesian approach that estimates the population parameters with a Bayesian paradigm. As in all Bayesian problems, it requires the definition of a prior on the population parameters $\boldsymbol{\Pi}$ (to which may be added $\boldsymbol{\omega}$, thus getting $\boldsymbol{\Psi}$) and it enables to derive the expression of the density of the population parameters given the available data thanks to the Bayes rule (also known as the posterior distribution). To sample the posterior distribution, a strategy based on Markov-Chain-Monte-Carlo (MCMC) algorithms should be implemented, which makes the calibration process expensive and difficult to carry out. An easier alternative consists in providing a frequentist estimation of the (full) population parameters, with, if necessary, a quantification of the uncertainties on the (full) population parameters (that is provided by the posterior distribution from the HBMs). One popular method to perform this calibration are the mixed-effects models [Demidenko, 2013, Lavielle, 2014, Pinheiro and Bates, 2002]. Both alternatives are described in the sections thereafter before being compared.

2.2.3 Hierarchical Bayesian models

The population modeling exhibits a hierarchical structure as illustrated in Figure 2.2, which shows that the individual level is embedded within the population level. In fact, from a mathematical point of view, the hierarchical link between the two levels is represented by Eq. (2.6a). This hierarchical organization can be treated within the frame of Hierarchical Bayesian models (HBMs) that aim to handle situations where hierarchical relations arise [Kruschke, 2014, Rouder and Lu, 2005]. They were introduced in behavior sciences [Gill, 2007], where researchers aimed to process data from different groups within the same statistical model. Applied in our context, these groups refer to the experiments, the lowest level refers to the individual level and the next higher to the population level. HBMs estimate the distribution of the population parameters (the top level variables) but also the distribution of the lowest level variables (the individual parameters). More formally, the observed output variables for each specimen $\mathbf{y}_i$ can be mimicked by a behavior law $\mathcal{F}(\mathbf{t}_i, \cdot)$, parameterized by the lowest level variables (in our case the individual parameters $\boldsymbol{\theta}_i$s) that are supposed to be drawn from a higher level distribution (that would be here $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ parameterized by $\boldsymbol{\Pi}$) to which is added an error term $\boldsymbol{\xi}_i$ drawn from $\mathcal{N}(\mathbf{0}_{\mathbb{R}_i^N}, \omega_i^2 \mathbf{I}_{N_i})$, and the hyper-parameters $\boldsymbol{\Psi} = [\boldsymbol{\Pi}, \boldsymbol{\omega}]$ are themselves sampled from another distribution $f_{\boldsymbol{\Psi}|\boldsymbol{\gamma}}$ parameterized by new hyper-parameters $\boldsymbol{\gamma}$, that can be themselves sampled from another distribution $f_{\boldsymbol{\gamma}}$, parameterized by new hyper-parameters [Kruschke, 2014]. If there were more than two levels, the same reasoning could be recursively continued. The statistical modeling of HBMs is the same as the one developed in Eqs. (2.6) to which is added a modeling of the variability of $\boldsymbol{\Psi}$ (*i.e.* $\boldsymbol{\Psi} \sim f_{\boldsymbol{\Psi}|\boldsymbol{\gamma}}$). This distribution corresponds to a prior distribution that is encountered in usual Bayesian analyses [Kennedy and O'Hagan, 2001, Wasserman, 2010]. Given this chain of dependencies, the posterior density of the population parameters can be expressed thanks to Bayes' rule as [Wakefield, 1996]:

$$\begin{aligned} f(\boldsymbol{\Psi}|\{\mathbf{y}_i\}_{i\in[\![1,n]\!]}, \boldsymbol{\gamma}) \propto & \quad f(\{\mathbf{y}_i\}_{i\in[\![1,n]\!]}|\{\boldsymbol{\theta}_i\}_{i\in[\![1,n]\!]}, \boldsymbol{\Pi}, \boldsymbol{\omega}, \boldsymbol{\gamma}) \ \} \text{individual level} \\ & \times \quad f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\{\boldsymbol{\theta}_i\}_{i\in[\![1,n]\!]}|\boldsymbol{\Pi}, \boldsymbol{\gamma}) \ \} \text{population level} \\ & \times \quad f(\boldsymbol{\Pi}|\boldsymbol{\gamma}) f(\boldsymbol{\omega}|\boldsymbol{\gamma}) \ \} \text{prior on the full population parameters} . \end{aligned} \tag{2.7}$$

Note that in Eq. (2.7) and for the remainder of the manuscript, the PDF of **all** densities **except** $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ will be noted with the generic letter f for the sake of clarity. Aside from the inherent variability of subjects [Rouder and Lu, 2005] and the measurements error, HBMs can embed if deemed necessary changes in the environmental conditions (*e.g.* thermal or moisture conditions

[Behmanesh et al., 2015, Song et al., 2020]). In particular, HBMs apply when the data consists in repetitions of the same test (this is our field of application) [Song et al., 2020]. A more general framework that accounts for uncertainties on the input data is available in [Nagel and Sudret, 2016]. It is possible to include as many levels as necessary in the modeling [Wu et al., 2018]. For each level, the distribution of the associated variables can be seen as a prior distribution and the immediate upper variables as the hyper-parameters of this distribution. Yet, in usual applications and as applied in Eq. (2.7), there are mainly 3 levels: the first could correspond to the individuals (Eqs. (2.6c) and (2.6b)), the second would describe the variability of the individual properties (Eq. (2.6a)) and the third would consist in a prior on the full population parameters: $\boldsymbol{\Psi} \sim f_{\boldsymbol{\Psi}}$ parameterized by known deterministic parameters $\boldsymbol{\gamma}$. This new hierarchical level could be illustrated in Figure 2.2 by adding another level that would encompass both the population and individual levels.
The HBMs approaches are fully Bayesian and use MCMC techniques to sample posterior distributions (*e.g.* the Metropolis-Hastings (M-H) algorithm [Wakefield, 1996] or the Gibbs sampler [Damlen et al., 1999]). Using MCMC techniques is computationally intensive and may not be feasible for applications involving computationally costly simulation codes. Futhermore, here, because of the hierarchical structure, the sampling gets more complicated than with the classical Bayesian techniques as the different levels depend on the upper one. Indeed, in Eq. (2.7), we can notice that the PDF of the first line depends on the value of $\boldsymbol{\theta}_i$, which is sampled according the second line, that depends on the value of $\boldsymbol{\Psi}$, that is drawn following the third line [Kruschke, 2014]. For instance, for a univariate linear model with 10 individuals, 25,000 samples are generated [Papadimas and Dodwell, 2021]. If the same number of samples is required for a more complex model that demands 1 second to be executed, the calibration would last at least 7 hours. Thus, for complex non-linear behavior laws that takes longer to run (and usually have many parameters to estimate), it does not seem possible to apply HBMs. The computational costs could be decreased with the use of surrogate models [Forrester et al., 2008], but they are complex to set up with models in high dimensions that describe highly non-linear behaviors. This complexity also concerns the diagnosis of convergence of the MCMC sampling [Geweke, 1989]. The choice of the prior, a central notion in Bayesian analysis, gets harder as it concerns the population parameters (if the individual parameters are Gaussian distributed, a usual prior for the covariance matrix $\boldsymbol{\Sigma}$ is an inverse-Wishart distribution [Schuurman et al., 2016]). HBMs are commonly used in many fields such as social and behavorial sciences [Gill, 2007], and they have already been applied in mechanics, in particular in structural mechanics [Ballesteros et al., 2014, Behmanesh et al., 2015, Jiang et al., 2010, Song et al., 2020]. These applications were motivated by the variability of the properties of concrete with respect to the environmental conditions.

2.2.4 Mixed-effects methods

Though a Bayesian calibration of the full population parameters may seem appealing as it provides their posterior distribution from which many different analyses can be performed, the computational resources mobilized and the duration of the calibration can make it infeasible in practice. Nevertheless, even if it is less informative, a point estimate of the population parameters can give a first insight into material variability. To estimate the population parameters in a frequentist framework, different approaches can be set up. Among those can be found (non-linear)mixed-effects models [Lavielle, 2014], the two-stage approach (TS) [Beal and Sheiner, 1982] and the naive-pooled data approach (NPD) [Sheiner and Beal, 1981]. The NPD assumes that all specimens are characterized by the same set of model parameters, the fixed-effects. It refers actually to the individual approach. The NPD method consists in pooling the data from all individuals together. To estimate the population averaged behavior, the sum of squares of the residuals for all individuals and observations is minimized (which is tantamount to maximizing the product of the marginal likelihoods for all individuals). The inter-individual

variability can be defined as the 95% confidence region, estimated either with the asymptotic theory or bootstrap as described in Chapter 1. Yet, as with multi-objective techniques, this estimation mixes all types of uncertainties together and therefore does not address the problems encountered in the first chapter. On the contrary, in the TS approach, the individuals do not share common parameters. The TS approach is decomposed in two different steps. First, the individual parameters are estimated separately minimizing classical least squares criterion for each specimen. Then, the model parameters distribution are estimated by maximum likelihood [Demidenko, 2013]. Though the TS approach can provide more consistent results than the NPD approach [Beal and Sheiner, 1982], it provides biased estimation of the population parameters, especially for those that characterize the inter-individual variability. For instance, they can be over-estimated because the individual parameters are estimated without consideration of the rest of the population [Beal and Sheiner, 1982].
Mixed-effects models [Demidenko, 2013, Lavielle, 2014, Pinheiro and Bates, 2002] provide other ways to estimate the full population parameters. As mentioned earlier, in the population approach, the variability of the individual parameters is modeled by a probability distribution. The combination of both fixed and random-effects into the same model can lead to mixed-effects models [Pinheiro and Bates, 2002]. Mixed-effects models can be compared to a combination of Analysis of Variance (Anova) models [Girden, 1992] (individual variability) and regression analysis [Demidenko, 2013]. They are relevant to implement when the data to process comes from repeated measurements or can be grouped into different categories. They have been applied extensively to calibrate linear/generalized linear model [Henderson et al., 1959, Laird and Ware, 1982] and non-linear models [Davidian and Giltinan, 1995, Pinheiro and Bates, 2002] in many different fields: pharmacometrics [Drikvandi, 2017, Ette and Williams, 2004], economics [Nelson and Siegel, 1987, Svensson, 1994], agriculture [Baey et al., 2016, Hall and Clutter, 2004], oil & gas industry [Harris and Lee, 2014, Lewis and Beal, 1918], epidemiology [Wang et al., 2020] among others. They are also suited for the analysis of clustered data [Demidenko, 2013].

The key assumption of mixed-effects is the decomposition of the individual parameters $\boldsymbol{\theta}_i$ through fixed-effects $\boldsymbol{\beta}$ and random-effects $\mathbf{b}_i$ as:

$$\boldsymbol{\theta}_i = h(\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{A}_i, \mathbf{B}_i) \ , \tag{2.8}$$

with $h(\cdot)$ a (non-linear) function, and $\mathbf{A}_i, \mathbf{B}_i$ deterministic design matrices that belong to $\mathcal{M}_d(\mathbb{R})$ with d the number of model parameters. For a linear function $h(\cdot)$, Eq. (2.8) can be transformed into:

$$\boldsymbol{\theta}_i = \mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i \ . \tag{2.9}$$

Design matrices other than $\mathbf{I}_d$, the identity matrix of $\mathcal{M}_d(\mathbb{R})$, can be used if some of the model parameters are deterministic to cancel their associated random-effects [Pinheiro and Bates, 2002]. For the sake of simplicity, $h(\cdot)$ will be chosen as a linear function for the remainder of this manuscript. Non-linear functions are useful to implement properties on the individual parameters (positivity for instance) but also hardens the calibration process. Furthermore, given that all parameters will be considered as random, the design matrices are set equal to $\mathbf{I}_d$. If other choices are made (with a non-linear function $h(\cdot)$ for instance), the derivations should be updated by applying the appropriate change of variables. Note that there is not necessarily a direct interpretation of the fixed-effects and random-effects with respect to the population parameters. Yet, for a Gaussian distribution, it is possible to decompose the individual parameters as:

$$\boldsymbol{\theta}_i = \boldsymbol{\mu} + \mathbf{b}_i \ . \tag{2.10}$$

with $\mathbf{b}_i \underset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_{\mathbb{R}^d}, \boldsymbol{\Sigma})$ with $\mathbf{0}_{\mathbb{R}^d}$ the vector of zeros in $\mathbb{R}^d$. Comparing Eqs. (2.10) and (2.9), it follows that the individual parameters are decomposed in a straightforward fashion between

fixed-effects $\boldsymbol{\beta} = \boldsymbol{\mu}$ and random-effects $\mathbf{b}_i$, with design matrices equal to $\mathbf{I}_d$ and a linear function $h(\cdot)$.
Mixed-effects modeling is compatible with most choices of probability distribution and types of residuals, including non-parametric distributions [Mallet et al., 1998]. Yet, most of the time, in population-based calibration [Pinheiro and Bates, 2002], the model parameters distribution is a Gaussian distribution (of dimension d) characterized by a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$. Because they are not bounded, Gaussian distributions are not completely satisfactory in problems where the parameters are bounded (even if they can be bounded to exclude non-physical values). However, the probabilities of exceeding the bounds are small and the Gaussian assumption helps in keeping the calculations tractable as it will be shown later in Section 2.3.2. This can be also justified by the maximum entropy theorem: Gaussian distributions exhibit the highest entropy (they are the least informative in other words) among other distributions given constraints on their second-order moment [Jaynes, 1957]. It gives:

$$f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}} = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \ , \tag{2.11}$$

with $\boldsymbol{\mu} \in \mathbb{R}^d$ the mean vector and $\boldsymbol{\Sigma} \in \mathcal{M}_d(\mathbb{R})$ the covariance matrix.

Mixed-effects can be generalized to several levels of grouping [Pinheiro and Bates, 2002]. In mechanics, grouping criteria could refer to the type of experiments (shear or bending test for instance) or to the nature of the load profile (traction or compression). For two levels of modeling, the k^{th} observation of the j^{th} second-level group of the i^{th} first-level group y_{ijk} reads as:

$$y_{ijk} = \mathcal{F}(t_{ijk}, \boldsymbol{\theta}_{ij}) + \xi_{ijk} \ \forall (i,j,k) \in [\![1,n]\!] \times [\![1,N_i]\!] \times [\![1,\tilde{N}_{ij}]\!] \ , \tag{2.12}$$

with $\tilde{N}_{ij}$ the number of observations of the j^{th} second-level group of the i^{th} first-level group and k the corresponding index, N_i the number of second-level groups of the i^{th} first-level group with j the corresponding index and n the number of first-level groups and i the corresponding index. For Gaussian distributed individual parameters, the $\boldsymbol{\theta}_i$s read as:

$$\boldsymbol{\theta}_{ij} = \boldsymbol{\mu} + \mathbf{b}_i + \mathbf{b}_{ij} \ , \tag{2.13}$$

with $\mathbf{b}_i$ the first-level random-effects and $\mathbf{b}_{ij}$ the second-level random-effects. These random-effects comply with $\mathbf{b}_i \underset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_{\mathbb{R}^d}, \boldsymbol{\Sigma}_1)$ and $\mathbf{b}_{ij} \underset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_{\mathbb{R}^d}, \boldsymbol{\Sigma}_2)$. In Eq. (2.13), $\mathbf{b}_i$ represent the between-group random-effects and $\mathbf{b}_{ij}$ the within-group random-effects. Similarly, $\boldsymbol{\Sigma}_1$ represents the between-group covariance matrix and $\boldsymbol{\Sigma}_2$ represents the within-group covariance matrix. The parameters $\boldsymbol{\Pi}$ to estimate here refer to $\boldsymbol{\Pi} = [\boldsymbol{\mu}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2]$. As in Eq. (2.9), design matrices can be added to the different terms of Eq. (2.13).

2.2.5 Links between HBMs and mixed-effects

Mixed-effects can be applied either in a frequentist or a Bayesian fashion [Lee, 2022]. In fact, when applied to characterize an intrinsic variability that arises through repetitions of specimens, HBMs turn out to be equivalent to Bayesian mixed-effects [Damlen et al., 1999, Pinheiro and Bates, 2002, Qiu et al., 2002, Wakefield, 1996]. Indeed, it appears that both of these methods exhibit strong similarities. Both of them *de facto* propose the same hierarchical modeling, even if the HBMs seem to be slighlty more general for other implementations. Both methods can handle several level of modelings (this point is detailed below for mixed-effects), even if this necessarily overburdens the computational costs [Wu et al., 2018]. Yet, it is interesting to notice that HBMs methods are only applicable in a Bayesian framework while mixed-effects methods can employed in both frequentist and Bayesian fashions [Lee, 2022]. This provides more flexibility about the type of resolution techniques that is chosen to determine the full

population parameters and their associated distribution. This choice is made according to the available computational facilities and model complexity among other criteria. To analyze the consequences of the hierarchical modeling applied to the calibration of behavior laws dedicated to composites material, we choose to focus on mixed-effects techniques, advancing to HBMs if the hierarchical modeling proves to be appropriate.

The statistical formalization presented in this section provides tools to model many situations. Given this modeling, it is possible to go through the joint estimation of the population parameters and residuals models ($\boldsymbol{\Psi}$), which is carried out by maximum likelihood in the frequentist framework.

2.3 Estimating the model parameters distribution

The frequentist estimation of the full population parameters is based on a specific likelihood that is derived in the next sections. Multivariate integrals are found in its derivation. Thus, to conduct likelihood maximization, two main trends can be identified. The first refers to a direct likelihood maximization, demanding to approximate it (*e.g.* with quadrature techniques [Vonesh et al., 2002], first-order methods [Beal and Sheiner, 1982, Demidenko, 2013], Laplace approximation [Demidenko, 2013, Pinheiro and Bates, 2002], (MCMC-)importance sampling [Lavielle, 2014, Pinheiro and Bates, 1994]). Another way to achieve calibration with mixed-effects relies on iterative procedures that alternate between approximations and maximizations of the likelihood as proposed by the Lindstrom & Bates algorithm [Lindstrom and Bates, 1990] or with the Stochastic Approximation Expectation Maximization (SAEM) [Kuhn and Lavielle, 2004]. All derivations assume a linear relation with the identity matrix as design matrices between the individuals parameters, and with the fixed and random-effects as in Eq. (2.10). Such a formulation can extended to the one presented in Eq. (2.9) *via* appropriate change of variables.

2.3.1 Likelihood of the mixed-effects

The estimation of full population parameters with mixed-effects for non-linear models is mostly performed with maximum likelihood (ML) or with methods related to it [Beal and Sheiner, 1982, Lavielle, 2014, Pinheiro and Bates, 2002]. This section details the derivation of the likelihood function for mixed-effects problems.

The likelihood refers to $\mathcal{L}(\boldsymbol{\Psi}) \coloneqq f(\mathbf{y}_1, \dots, \mathbf{y}_n|\boldsymbol{\Psi})$, *i.e.* the joint PDF of the output measures $\{\mathbf{y}_i\}_{i\in[\![1,n]\!]}$ given the full population parameters $\boldsymbol{\Psi}$. Following the assumptions made in Eqs (2.6), the different individuals can be considered as independent. Then, the likelihood reads as the product of all the individual likelihoods, $\mathcal{L}_i(\boldsymbol{\Psi}) \coloneqq f(\mathbf{y}_i|\boldsymbol{\Psi})$,

$$\mathcal{L}(\boldsymbol{\Psi}) \coloneqq \prod_{i=1}^{n} \mathcal{L}_i(\boldsymbol{\Psi}) \ . \tag{2.14}$$

Because the $\boldsymbol{\theta}_i$s are not observed, the likelihood of the i^{th} individual $\mathcal{L}_i(\boldsymbol{\Psi})$ is the integral of $f(\mathbf{y}_i, \boldsymbol{\theta}_i|\boldsymbol{\Pi}, \omega_i)$ (*i.e.* the density of the output data $\mathbf{y}_i$ given individual parameters $\boldsymbol{\theta}_i$ and the full population parameters $\boldsymbol{\Psi}$) with respect to all possible $\boldsymbol{\theta}_i$ over $\mathbb{R}^d$:

$$\mathcal{L}_i(\boldsymbol{\Psi}) \coloneqq \int_{\mathbb{R}^d} f(\mathbf{y}_i, \boldsymbol{\theta}_i|\boldsymbol{\Pi}, \omega_i) \mathrm{d}\boldsymbol{\theta}_i \ . \tag{2.15}$$

Following the conditional probability rule, we have:

$$f(\mathbf{y}_i, \boldsymbol{\theta}_i|\boldsymbol{\Pi}, \omega_i) = f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) \ . \tag{2.16}$$

In Eq. (2.16), $f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i)$ refers to the probability to observe output $\mathbf{y}_i$, given the model output for the individual parameter $\boldsymbol{\theta}_i$, itself drawn from $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})$, the density of the individual parameters $\boldsymbol{\theta}_i$ given the full population parameters $\boldsymbol{\Psi}$ (*i.e.* the joint PDF of the model parameters distribution) and the measurements errors $\boldsymbol{\xi}_i$ characterized in Eq. (2.6c). In Eq. (2.16), $f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i)$ is a fitting term (as for classical approaches) and $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})$ stands for a natural regularization term brought by the mixed-effects framework. Yet, this regularization is not arbitrary as with Tikhonov formulation [Tikhonov et al., 1995] but arises from the assumption on $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ [Demidenko, 2013]. The individual likelihoods can be further transformed into:

$$\mathcal{L}_i(\boldsymbol{\Psi}) = \int_{\mathbb{R}^d} f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) \mathrm{d}\boldsymbol{\theta}_i = \int_{\mathbb{R}^d} \exp(-g_i(\mathbf{y}_i, \boldsymbol{\theta}_i, \boldsymbol{\Psi})) \mathrm{d}\boldsymbol{\theta}_i \ , \tag{2.17}$$

with

$$g_i(\mathbf{y}_i, \boldsymbol{\theta}_i, \boldsymbol{\Psi}) \coloneqq -\ln(f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \boldsymbol{\omega}) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})) \ . \tag{2.18}$$

If the $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ is a multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $f(\mathbf{y}_i, \boldsymbol{\theta}_i|\boldsymbol{\Psi})$ reads as:

$$f(\mathbf{y}_i, \boldsymbol{\theta}_i|\boldsymbol{\Pi}, \omega_i) = \frac{1}{\omega_i^{N_i}\sqrt{|\boldsymbol{\Sigma}|(2\pi)^{N_i+d}}} \exp\left(-\tilde{g}_i(\mathbf{y}_i, \boldsymbol{\mu} + \mathbf{b}_i, \boldsymbol{\Delta}_i)\right) \ . \tag{2.19}$$

with function $\tilde{g}_i(\cdot)$ defined as:

$$\tilde{g}_i(\mathbf{y}_i, \boldsymbol{\mu} + \mathbf{b}_i, \boldsymbol{\Delta}_i) = \frac{\|\mathbf{y}_i - \mathcal{F}(\mathbf{t}_i, \boldsymbol{\mu} + \mathbf{b}_i)\|^2 + \|\boldsymbol{\Delta}_i \mathbf{b}_i\|^2}{2\omega_i^2} \ , \tag{2.20}$$

where $\boldsymbol{\Delta}_i$ is the transpose of the Cholesky decomposition of $\omega_i^2\boldsymbol{\Sigma}^{-1}$ (so $\omega_i^2\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Delta}_i^\top\boldsymbol{\Delta}_i$) after applying the change of variable defined in Eq. (2.10). Note that functions $g_i(\cdot)$ defined in Eq. (2.18) and $\tilde{g}_i(\cdot)$ in Eq. (2.20) match if the model parameters distribution is multivariate Gaussian and the residuals are Gaussian distributed, up to the constant term with respect to $\boldsymbol{\theta}_i$, $\frac{1}{\omega_i^{N_i}\sqrt{|\boldsymbol{\Sigma}|(2\pi)^{N_i+d}}}$. The individual likelihoods can be further detailed into [Pinheiro and Bates, 2002]:

$$\begin{aligned}\mathcal{L}_i(\boldsymbol{\Psi}) &= \int_{\mathbb{R}^d} f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) \mathrm{d}\boldsymbol{\theta}_i \\ &= \int_{\mathbb{R}^d} \frac{1}{(\omega_i\sqrt{2\pi})^{N_i}\sqrt{|\boldsymbol{\Sigma}|(2\pi^d)}} \exp\left(-\tilde{g}_i(\mathbf{y}_i, \boldsymbol{\mu} + \mathbf{b}_i, \boldsymbol{\Delta}_i)\right) \mathrm{d}\mathbf{b}_i \ .\end{aligned} \tag{2.21}$$

For a linear model, *i.e.* $\mathcal{F}(\mathbf{t}_i, \boldsymbol{\theta}_i) = \mathbf{R}(\mathbf{t}_i)\boldsymbol{\theta}_i = \mathbf{X}(\mathbf{t}_i)\boldsymbol{\mu} + \mathbf{Z}(\mathbf{t}_i)\mathbf{b}_i$ with $\mathbf{R}(\mathbf{t}_i)$ the matrix of predictors, $\mathbf{X}(\mathbf{t}_i) = \mathbf{R}(\mathbf{t}_i)\mathbf{A}_i$ and $\mathbf{Z}(\mathbf{t}_i) = \mathbf{R}(\mathbf{t}_i)\mathbf{B}_i$ that belong to $\mathcal{M}_{N_i d}(\mathbb{R})$, the likelihood of linear mixed-effects models can be computed in closed form [Pinheiro and Bates, 2002] and is expressed as:

$$\mathcal{L}_{\mathrm{LME}}(\boldsymbol{\Psi}) = \prod_{i=1}^{n} \frac{|\boldsymbol{\Delta}_i| \exp\left(-\tilde{g}_i(\mathbf{y}_i, \boldsymbol{\mu} + \widehat{\mathbf{b}}_i, \boldsymbol{\Delta}_i)\right)}{\sqrt{(2\pi)^{N_i}|\mathbf{Z}(\mathbf{t}_i)^\top\mathbf{Z}(\mathbf{t}_i) + \boldsymbol{\Delta}_i^\top\boldsymbol{\Delta}_i|}} \ , \tag{2.22}$$

with $\widehat{\mathbf{b}}_i$ the estimated individual deviations that stand for $\widehat{\mathbf{b}}_i = (\mathbf{Z}(\mathbf{t}_i)^\top\mathbf{Z}(\mathbf{t}_i) + \boldsymbol{\Delta}_i^\top\boldsymbol{\Delta}_i)^{-1}(\mathbf{Z}(\mathbf{t}_i) + \boldsymbol{\Delta}_i)^\top(\mathbf{y}_i - \mathbf{X}(\mathbf{t}_i)\boldsymbol{\mu})$.
Yet, for non-linear models or non Gaussian distributions, the exact expression of the likelihood no longer holds and the likelihood should be approximated. The next section describes the usual ways with mixed-effects models to perform this estimation.

2.3.2 Likelihood estimation

To proceed to the direct maximization of the likelihood, it is necessary to estimate it. Remember that it is expressed as a product of multivariate integrals that stand for:

$$\mathcal{L}(\boldsymbol{\Psi}) = \prod_{i=1}^{n} \int_{\mathbb{R}^d} f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) \mathrm{d}\boldsymbol{\theta}_i \ . \tag{2.23}$$

Unlike in the specific case described in the previous section, there is no analytical expression of the multivariate integral and its computation is challenging as it requires approximating integrals. Usual methods to estimate integrals (*e.g.*, quadrature, sampling-based techniques among which the Monte-Carlo method) are presented. For weakly non-linear models, linear approximations of the behavior law can employed, which enables to use the expression of the likelihood provided in Eq. (2.22). More generally, if the function $g_i(\cdot)$ (not only the behavior law) complies with regularity conditions, a Laplace approximation of the likelihood can be preferred instead.

2.3.2.1 First order methods [Beal and Sheiner, 1982]

First order methods aim to take advantage of the fact that the likelihood can be analytically computed for a linear model, considering Gaussian distributed residuals and Gaussian distributed model parameters. They are based on a Taylor expansion around the fixed-effects $\boldsymbol{\mu}$ of the model as follows:

$$\mathbf{y}_i = \mathcal{F}(\mathbf{t}_i, \boldsymbol{\theta}_i) + \boldsymbol{\xi}_i \approx \mathcal{F}(\mathbf{t}_i, \boldsymbol{\mu}) + \frac{\partial \mathcal{F}(\mathbf{t}_i, \boldsymbol{\mu})}{\partial \mathbf{b}_i} \mathbf{b}_i + \boldsymbol{\xi}_i \ . \tag{2.24}$$

The likelihood can be then computed in closed form (see Eq. (2.22)), allowing its maximization. An obvious drawback of this method is that it is based on a strong approximation of the model. Indeed, the Taylor expansion is performed around the fixed-effects and not around the random-effects, which can harm the interest of the method if the random-effects are located far from the fixed-effects. Moreover, it only allows the choice of a multivariate normal distribution for $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ because otherwise, the exact calculation is not available.

2.3.2.2 Laplace approximation [Pinheiro and Bates, 2002, Wolfinger, 1993]

The Laplace approximation applies to integrals of the type

$$A = \int_{\mathbb{R}^d} e^{-w(\mathbf{x})} \mathrm{d}\mathbf{x} \ , \tag{2.25}$$

with $w(\cdot)$ a function which complies with the following constraints:

1. $w(\cdot)$ admits a global minimum $\mathbf{x}_0$ that belongs to the integration interval,
2. $w(\cdot)$ is a twice-differentiable function,
3. its Hessian matrix computed at $\mathbf{x} = \mathbf{x}_0$ is a symmetric definite positive matrix.

The main idea is to state that only points close to $\mathbf{x}_0$ significantly contribute to the integral. The different calculations that allow to establish the equations of the Laplace approximation are presented below. The Taylor expansion of $w(\cdot)$ at $\mathbf{x}_0$ can be written as:

$$w(\mathbf{x}) = w(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^\top \nabla w(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^\top \mathcal{H}(w)(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|^2) \ , \tag{2.26}$$

with:

- $\nabla w(\mathbf{x}_0) = (\frac{\partial w}{\partial \mathbf{x}_i}(\mathbf{x}_0))_{i\in[\![1,d]\!]}$ the gradient vector of $w(\cdot)$ at $\mathbf{x}_0$,
- $\mathcal{H}(w)(\mathbf{x}_0) = (\frac{\partial^2 w}{\partial \mathbf{x}_i \partial \mathbf{x}_j}(\mathbf{x}_0))_{(i,j)\in[\![1,d]\!]^2}$ the Hessian matrix of $w(\cdot)$ at $\mathbf{x}_0$.

As $\mathbf{x}_0$ is the global minimum, the gradient vanishes and the substitution of $w(\mathbf{x})$ in Eq. (2.25) by its approximation determined in Eq. (2.26) gives:

$$\begin{aligned}\int_{\mathbb{R}^d} e^{-w(\mathbf{x})}\mathrm{d}\mathbf{x} &\approx \int_{\mathbb{R}^d} e^{-w(\mathbf{x}_0)-\frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^\top \mathcal{H}(w)(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)}\mathrm{d}\mathbf{x} \\ &\approx e^{-w(\mathbf{x}_0)} \int_{\mathbb{R}^d} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^\top \mathcal{H}(w)(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)}\mathrm{d}\mathbf{x} \ .\end{aligned} \tag{2.27}$$

$\mathcal{H}(w)(\mathbf{x}_0)$ is symmetric positive-definite, as a result, in Eq. (2.27), the integrand is a Gaussian PDF of mean $\mathbf{x}_0$ and covariance matrix $\mathcal{H}(w)(\mathbf{x}_0)^{-1}$. As a PDF always integrates to 1,

$$\int_{\mathbb{R}^d} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^\top M^{-1}(\mathbf{x}-\mathbf{x}_0)}\mathrm{d}\mathbf{x} = (2\pi)^{\frac{d}{2}}\sqrt{|M|} \ . \tag{2.28}$$

Applying Eqs. (2.27) and (2.28), the Laplace approximation of A is obtained as

$$A \approx e^{-w(\mathbf{x}_0)} \frac{(2\pi)^{\frac{d}{2}}}{\sqrt{|\mathcal{H}(w)(\mathbf{x}_0)|}} \ . \tag{2.29}$$

In model calibration, the value of the global minimum is unknown. Thus, to apply this method, it is necessary to determine this extremum that stands for the individual parameters $\boldsymbol{\theta}_i$ estimated for each individual by minimizing $g_i(\cdot)$ from Eq. (2.18), that is to say the solution of

$$\widehat{\boldsymbol{\theta}}_i = \underset{\boldsymbol{\theta}_i\in\Gamma_{\boldsymbol{\theta}}}{\arg\min}\, g_i(\mathbf{y}_i, \boldsymbol{\theta}_i, \boldsymbol{\Psi}) \ , \tag{2.30}$$

with $\Gamma_{\boldsymbol{\theta}}$ the set admissible model parameters. Then, Eq. (2.29) can be used to compute the Laplace approximation. The negative log-likelihood finally is expressed as:

$$\begin{aligned}-\ln(\mathcal{L}(\boldsymbol{\Psi})) &= -\sum_{i=1}^{n} \ln\left(\mathcal{L}_i(\boldsymbol{\Psi})\right) = -\sum_{i=1}^{n} \ln\left(\int_{\mathbb{R}^d} f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})\mathrm{d}\boldsymbol{\theta}_i\right) \\ &\approx \sum_{i=1}^{n}\left(\frac{1}{2}\ln(|\mathcal{H}(g_i)(\widehat{\boldsymbol{\theta}}_i)|) + g_i(\mathbf{y}_i, \widehat{\boldsymbol{\theta}}_i, \boldsymbol{\Psi})\right) - \frac{d}{2}\ln(2\pi) \ .\end{aligned} \tag{2.31}$$

For Gaussian distributed residuals, the Hessian matrix of $g_i(\cdot)$ at $\widehat{\boldsymbol{\theta}}_i$ is expressed as [Pinheiro and Bates, 2002]:

$$\begin{aligned}\mathcal{H}(g_i)(\widehat{\boldsymbol{\theta}}_i) = \frac{1}{\omega_i^2}\left(\frac{\partial^2 \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_i^\top}(\mathbf{y}_i - \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)) + \frac{\partial \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i}^\top \frac{\partial \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i}\right) \\ - \frac{\partial^2 \ln(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\widehat{\boldsymbol{\theta}}_i|\boldsymbol{\Psi}))}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_i^\top} \ .\end{aligned} \tag{2.32}$$

In practice, $\frac{\partial^2 \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_i^\top}(\mathbf{y}_i - \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i))$ can be neglected if the model $\mathcal{F}(\mathbf{t}_i, \boldsymbol{\theta}_i)$ is close enough to the experiment $\mathbf{y}_i$ [Bates et al., 1983]. The term $\frac{\partial \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i}$ is evaluated with a finite difference scheme and the accuracy of the estimation checked before proceeding to the final computation. For Gaussian distributed parameters, $\frac{\partial^2 \ln(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\widehat{\boldsymbol{\theta}}_i|\boldsymbol{\Psi}))}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_i^\top}$ reduces to $-\boldsymbol{\Sigma}^{-1}$. Differentiating function $\tilde{g}_i(\cdot)$ either with respect to $\boldsymbol{\theta}_i$ or $\mathbf{b}_i$ gives the same result given the chain rule following Eq.

(2.10). The log-likelihood can be expressed as a function of both variables. It can be further detailed into:

$$\begin{aligned} -\ln(\mathcal{L}(\boldsymbol{\Psi})) &= -\sum_{i=1}^{n} \ln\left(\mathcal{L}_i(\boldsymbol{\Psi})\right) = -\sum_{i=1}^{n} \ln\left(\int_{\mathbb{R}^d} f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) \mathrm{d}\boldsymbol{\theta}_i\right) \\ &\approx \frac{n}{2}\ln(|\boldsymbol{\Sigma}|) + \sum_{i=1}^{n}\left(\frac{1}{2}\ln(|\mathcal{H}(\tilde{g}_i)(\widehat{\mathbf{b}}_i)|) + \frac{\tilde{g}_i(\mathbf{y}_i, \boldsymbol{\mu} + \widehat{\mathbf{b}}_i, \boldsymbol{\Delta}_i)}{2\omega_i^2}\right. \\ &\left. + N_i \ln(\omega_i\sqrt{2\pi})\right) \ , \end{aligned} \tag{2.33}$$

This expression is similar to the likelihood defined in Eq. (2.22) up to the design matrices $\mathbf{A}_i$ and $\mathbf{B}_i$, except that the individual deviations cannot be expressed in closed form. The main difference between the Laplace approximation and the first-order methods is that the Taylor expansion is carried out around the individual parameters with the Laplace method while it is achieved around the fixed-effects for the first-order methods. This is a clear advantage for the Laplace method, especially when the individual parameters $\boldsymbol{\theta}_i$ are located far from the fixed-effects, in other words when the Taylor expansion around the fixed-effects is no longer correct to estimate the experimental output of each individual $\mathbf{y}_i$.
A key feature of the Laplace approach is the resolution of the minimization problem defined in Eq. (2.30). Indeed, the estimation of the likelihood strongly relies on a proper resolution of this minimization problem that has to be performed accurately. In the SAS [SAS, 2004] and NONMEM software [Bauer, 2019], a Quasi-Newton method is chosen to perform this minimization. Another numerical issue stands in the determinant of $\mathcal{H}(g_i)(\widehat{\mathbf{b}}_i)$, that has to remain (strictly) positive to allow the computation. Finally, this approximation gets true when N_i is large enough, but remains valid in most cases [Davidian and Giltinan, 1995]. An order of magnitude of the number of observation points with respect to the number of parameters is proposed in [Ogden, 2021].

2.3.2.3 MCMC-IS [Geweke, 1989, Kuhn and Lavielle, 2004, Lavielle, 2014]

A fundamental method to compute integrals is the Monte-Carlo method [Metropolis and Ulam, 1949]. The direct application of the Monte-Carlo rule to approximate the global likelihood gives:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\Psi}) &= \prod_{i=1}^{n} \int_{\mathbb{R}^d} f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) \mathrm{d}\boldsymbol{\theta}_i \\ &\approx \prod_{i=1}^{n} \frac{1}{N_{MC}} \sum_{\ell=1}^{N_{MC}} f(\mathbf{y}_i|\boldsymbol{\theta}_{i,\ell}, \boldsymbol{\Pi}, \omega_i) \ , \end{aligned} \tag{2.34}$$

with $\{\boldsymbol{\theta}_{i,\ell}\}_{\ell\in[\![1,N_{MC}]\!]}$ i.i.d samples drawn from $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})$ for all individuals. One of the issues raised by Eq. (2.34) is that if one of the integrals is close to zero, the entire estimate goes down because of the product, thus flawing the estimation of the likelihood. Given that the sampling density $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})$ does not necessarily generate model parameters that result in a proper adequation of the model responses and the observations, the likelihood function often collapses to 0. One possibility to avoid this problem would to generate a huge number of samples but this is prohibitive from a computational point of view.
In order to generate model parameters that better suit the observations, an importance sampling scheme (written IS) may be set up [Lavielle, 2014, Pinheiro and Bates, 1994]. Importance sampling scheme belongs to the same family of integration schemes as the classical Monte-Carlo [Wasserman, 2010]. An auxiliary PDF is used to generate samples in place of the initial density function. The idea is to generate model parameters associated to the model responses that are more consistent with the available observations. The auxiliary distribution allows to generate

samples that are relevant to the proper estimation of the integral and as a result, the likelihood function does not collapse to 0, a necessary condition for its maximization. The first step consists in rewriting the integral of Eq. (2.15) as follows:

$$\int_{\mathbb{R}^d} f(\mathbf{y}_i|\boldsymbol{\theta}_i,\boldsymbol{\Pi},\omega_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})\mathrm{d}\boldsymbol{\theta}_i = \int_{\mathbb{R}^d} f(\mathbf{y}_i|\boldsymbol{\theta}_i,\boldsymbol{\Pi},\omega_i)\frac{f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})}{\pi_i(\boldsymbol{\theta}_i|\boldsymbol{\Psi})}\pi_i(\boldsymbol{\theta}_i|\boldsymbol{\Psi})\mathrm{d}\boldsymbol{\theta}_i \ , \tag{2.35}$$

with $\pi_i(\boldsymbol{\theta}_i|\boldsymbol{\Psi})$ the importance sampling density. The choice of the importance sampling density is a key feature. Among the possible choices, the one that allows to reduce the most the variance of the likelihood estimator and proposed in [Lavielle, 2014] is [Wasserman, 2010]:

$$\pi_i(\boldsymbol{\theta}_i|\boldsymbol{\Psi}) = \frac{f(\mathbf{y}_i|\boldsymbol{\theta}_i,\boldsymbol{\Pi},\omega_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})}{f(\mathbf{y}_i|\boldsymbol{\Psi})} = \frac{f(\mathbf{y}_i,\boldsymbol{\theta}_i|\boldsymbol{\Pi},\omega_i)}{f(\mathbf{y}_i|\boldsymbol{\Psi})} = f(\boldsymbol{\theta}_i|\mathbf{y}_i,\boldsymbol{\Pi},\omega_i) \ . \tag{2.36}$$

This proposal of sampling density is interesting because it allows to generate model parameters conditioned on both $\boldsymbol{\Psi}$ and the available observations $\mathbf{y}_i$. Even if this density is unknown and its computation intractable (especially because of the non-linearity of the behavior law), it can be sampled with MCMC scheme (*e.g.*, the Metropolis-Hastings algorithm [Hastings, 1970, Metropolis et al., 1953]) (which remains cheaper than the nested MCMC involved for the HBMs). Indeed, the numerator in Eq. (2.36) is known from Eq. (2.6c) and the denominator is a constant with respect to $\boldsymbol{\theta}_i$. This method is denoted MCMC-IS. A remaining problem to compute the likelihoods from Eq. (2.35) is to provide, for each individual, an estimation of $f(\boldsymbol{\theta}_i|\mathbf{y}_i,\boldsymbol{\Pi},\omega_i)$ that can be solved using Kernel Density Estimation methods [Wand and Jones, 1994]. However, with such techniques, the accuracy of the PDF estimation decreases when the dimension of $\boldsymbol{\theta}_i$ increases. In [Lavielle, 2014], it is approximated as:

$$\pi_i(\boldsymbol{\theta}_i|\boldsymbol{\Psi}) = m(\boldsymbol{\theta}_i|\mathbf{y}_i\boldsymbol{\Pi},\omega_i) + C(\boldsymbol{\theta}_i|\mathbf{y}_i,\boldsymbol{\Pi},\omega_i)T_\nu \ , \tag{2.37}$$

with $m(f(\boldsymbol{\theta}_i|\mathbf{y}_i,\boldsymbol{\Pi},\omega_i)) = \mathbb{E}(f(\boldsymbol{\theta}_i|\mathbf{y}_i,\boldsymbol{\Pi},\omega_i))$ the empirical mean of the MCMC samples of $f(\boldsymbol{\theta}_i|\mathbf{y}_i,\boldsymbol{\Pi},\omega_i)$ and $C^2(\boldsymbol{\theta}_i|\mathbf{y}_i,\boldsymbol{\Pi},\omega_i) = \mathbb{V}(\boldsymbol{\theta}_i|\mathbf{y}_i,\boldsymbol{\Pi},\omega_i)$, the empirical covariance matrix with only diagonal terms in $\mathcal{M}_d(\mathbb{R})$ and T_ν the Student's t-distribution with ν degrees of freedom. Values of appropriate degrees of freedom are proposed in [Lavielle, 2014].
The generation of MCMC samples at each step of the maximization is a computational burden which may be prohibitive for complex behavior laws (even if it is less computationally intensive as nested MCMCs involved in HMB). Instead of generating samples with MCMC methods, [Pinheiro and Bates, 1994] proposes another sampling density that naturally arises after a Taylor expansion of $\exp(-g_i(\mathbf{y}_i,\boldsymbol{\theta}_i,\boldsymbol{\Psi}))$ around $\widehat{\boldsymbol{\theta}}_i$, the solution of Eq. (2.30). It follows that:

$$\begin{aligned}\exp(-g_i(\mathbf{y}_i,\boldsymbol{\theta}_i,\boldsymbol{\Psi})) &\approx \exp\left(-\left(g_i(\mathbf{y}_i,\widehat{\boldsymbol{\theta}}_i,\boldsymbol{\Psi}) + \frac{1}{2}(\boldsymbol{\theta}_i-\widehat{\boldsymbol{\theta}}_i)^\top\mathcal{H}(g_i)(\widehat{\boldsymbol{\theta}}_i)(\boldsymbol{\theta}_i-\widehat{\boldsymbol{\theta}}_i)\right)\right) \\ &= \sqrt{\frac{(2\pi)^d}{|\mathcal{H}(g_i)(\widehat{\boldsymbol{\theta}}_i)|}}\exp\left(-g_i(\mathbf{y}_i,\widehat{\boldsymbol{\theta}}_i,\boldsymbol{\Psi})\right)\pi_i(\boldsymbol{\theta}_i|\boldsymbol{\Psi}) \ ,\end{aligned} \tag{2.38}$$

with $\pi_i(\boldsymbol{\theta}_i|\boldsymbol{\Psi}) \coloneqq \mathcal{N}(\widehat{\boldsymbol{\theta}}_i,\mathcal{H}(g_i)(\widehat{\boldsymbol{\theta}}_i)^{-1})$, with $\mathcal{H}(g_i)(\widehat{\boldsymbol{\theta}}_i)$ the Hessian matrix of $g_i(\cdot)$ at $\widehat{\boldsymbol{\theta}}_i$, the solution of the minimization problem of Eq. (2.30) (the gradient vanishes because the $\widehat{\boldsymbol{\theta}}_i$ is a minimum). Here, the term $f(\mathbf{y}_i|\boldsymbol{\theta}_i,\boldsymbol{\Pi},\omega_i)\frac{f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})}{\pi_i(\boldsymbol{\theta}_i|\boldsymbol{\Psi})}$ of Eq. (2.35) matches with $\sqrt{\frac{(2\pi)^d}{|\mathcal{H}(g_i)(\widehat{\boldsymbol{\theta}}_i)|}}$ $\times\exp\left(-g_i(\mathbf{y}_i,\widehat{\boldsymbol{\theta}}_i,\boldsymbol{\Psi})\right)$ in Eq. (2.38). This sampling density can be seen as an approximation by a Gaussian distribution of $f(\boldsymbol{\theta}_i|\mathbf{y}_i,\boldsymbol{\Pi},\omega_i)$ as:

$$f(\boldsymbol{\theta}_i|\mathbf{y}_i,\boldsymbol{\Pi},\omega_i) = \frac{f(\mathbf{y}_i,\boldsymbol{\theta}_i|\boldsymbol{\Pi},\omega_i)}{f(\mathbf{y}_i|\boldsymbol{\Psi})} \propto \exp(-g_i(\mathbf{y}_i,\boldsymbol{\theta}_i,\boldsymbol{\Psi})) \ , \tag{2.39}$$

Finally, for N_{IS} i.i.d. samples generated with respect to $\pi_i(\boldsymbol{\theta}_i|\boldsymbol{\Psi})$ (labeled $\{\boldsymbol{\theta}_{i,\ell}\}_{\ell\in[\![1,N_{IS}]\!]}$), the integral of Eq. (2.35) is approximated as follows:

$$\int_{\mathbb{R}^d} f(\mathbf{y}_i|\boldsymbol{\theta}_i,\boldsymbol{\Pi},\omega_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})\mathrm{d}\boldsymbol{\theta}_i \approx \frac{1}{N_{IS}}\sum_{\ell=1}^{N_{IS}} f(\mathbf{y}_i|\boldsymbol{\theta}_{i,\ell},\boldsymbol{\Pi},\omega_i)\frac{f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_{i,\ell}|\boldsymbol{\Pi})}{\pi_i(\boldsymbol{\theta}_{i,\ell}|\boldsymbol{\Psi})} \ . \tag{2.40}$$

Whichever sampling density is chosen, the computational cost of MCMC-IS to estimate the likelihood is large. This is the main drawback of this method. Its main advantage is that it applies to a wide range of assumptions of model parameters distributions. It also applies if the behavior law has poor regularity properties or a complicated numerical behavior, at the expense of a significant increase of model evaluations. In addition, even if the generation of samples is performed with the exact densities, it remains necessary to approximate it for the final estimation of the likelihoods, which implies another potential loss of accuracy for the likelihood maximization. The importance sampling density suggested by [Pinheiro and Bates, 1994] allows to partly address these problems by replacing the MCMC sampling by the resolution of a minimization problem and providing an exact evaluation of the importance weights. However, this solution only holds when function $g_i(\cdot)$ meets the same requirements as those prescribed for the Laplace approximation and is less general than the one implying MCMC sampling. Furthermore, contrary to the Laplace approximation, the assumption of positive-definite Hessian matrix is crucial. Indeed, in practice, the maximization with the Laplace approximation may proceed as long as its determinant is strictly positive, including non positive-definite matrices when the number of negative eigen-values is even. Here, if the inverse of the Hessian matrix has negative eigen-values, the maximization of the likelihood aborts as the sampling density is no more defined.

2.3.2.4 Quadrature

To estimate integrals, quadrature techniques can be employed. Remember that the likelihood is the integral of $f(\mathbf{y}_i|\boldsymbol{\theta}_i,\boldsymbol{\Pi},\omega_i)$ with respect to PDF $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})$. This makes Gaussian quadrature more attractive than its standard version as the corresponding weights and nodes incorporate the information of $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})$ [Pinheiro and Bates, 2002, Plan et al., 2012], especially when $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})$ refers to a known type of quadrature (Gauss-Hermite if $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})$ is a Gaussian multivariate distribution). In this case, noting $\alpha_i(\boldsymbol{\Psi}) \coloneqq \frac{1}{\omega_i^{N_i}\sqrt{|\boldsymbol{\Sigma}|(2\pi)^{N_i+d}}}$, the likelihood reads as:

$$\begin{aligned}
\mathcal{L}_i(\boldsymbol{\Psi}) &= \int_{\mathbb{R}^d} \alpha_i(\boldsymbol{\Psi}) \exp\left(-\left(\tilde{g}_i(\mathbf{y}_i,\boldsymbol{\theta}_i,\boldsymbol{\Psi})\right)\right)\mathrm{d}\boldsymbol{\theta}_i \\
&= \int_{\mathbb{R}^d} \alpha_i(\boldsymbol{\Psi}) \exp\left(-\left(\tilde{g}_i(\mathbf{y}_i,\mathbf{b}_i,\boldsymbol{\Delta}_i)\right)\right)\mathrm{d}\mathbf{b}_i \\
&= \int_{\mathbb{R}^d} \alpha_i(\boldsymbol{\Psi}) \exp\left(-\frac{\|\mathbf{y}_i-\mathcal{F}(\mathbf{t}_i,\boldsymbol{\mu}+\mathbf{b}_i)\|^2+\|\boldsymbol{\Delta}_i\mathbf{b}_i\|^2}{2\omega_i^2}\right)\mathrm{d}\mathbf{b}_i \\
&= \int_{\mathbb{R}^d} \frac{\alpha_i(\boldsymbol{\Psi})\omega_i^d}{|\boldsymbol{\Delta}_i|} \exp\left(-\frac{\|\mathbf{y}_i-\mathcal{F}(\mathbf{t}_i,\boldsymbol{\mu}+\omega_i^{-1}\boldsymbol{\Delta}_i\mathbf{z})\|^2}{2\omega_i^2}\right)\exp\left(-\frac{\|\mathbf{z}\|^2}{2}\right)\mathrm{d}\mathbf{z} \ .
\end{aligned} \tag{2.41}$$

In Eq. (2.41), variable $\mathbf{z}$ is defined by $\mathbf{z} \coloneqq (\omega_i)^{-1}\boldsymbol{\Delta}_i\mathbf{b}_i$ (and thus follows $\mathcal{N}(\mathbf{0}_{\mathbb{R}^d},\mathbf{I}_d)$). In the first line of Eq. (2.41), the change of variable corresponding to Eq. (2.10) is performed. In the last line, the change of variable that maps $\mathbf{b}_i$ into $\mathbf{z}$ is performed with Jacobian determinant $\frac{\omega_i^d}{|\boldsymbol{\Delta}_i|}$. The interest of variable $\mathbf{z}$ is that its coordinates are independent from one another. Therefore, with a tensor product of the quadrature rules in the different directions, the likelihood function can be estimated. With $z_\ell, w_\ell, \ell \in [\![1,N_{GQ}]\!]$ the weights and nodes corresponding to the univariate centered standardized Gaussian density, N_{GQ} the number of nodes and noting $\mathbf{z}_\ell = [z_{\ell_l}]_{l\in[\![1,d]\!]}$,

the individual likelihood approximates as:

$$\mathcal{L}_i(\boldsymbol{\Psi}) \approx \frac{\alpha_i(\boldsymbol{\Psi})\omega_i^d}{|\boldsymbol{\Delta}_i|} \sum_{\ell_1=1}^{N_{GQ}} \cdots \sum_{\ell_d=1}^{N_{GQ}} \exp\left(-\frac{\|\mathbf{y}_i - \mathcal{F}(\mathbf{t}_i, \boldsymbol{\mu} + \omega_i^{-1}\boldsymbol{\Delta}_i \mathbf{z}_\ell)\|^2}{2\omega_i^2}\right) \times \prod_{l=1}^{d} w_{\ell_l} \ . \tag{2.42}$$

Yet, given that the individual parameters that are the modes of PDF $f(\mathbf{y}_i, \boldsymbol{\theta}_i|\boldsymbol{\Psi})$ are unknown, direct quadrature requires to integrate over wide intervals and is usually replaced by adaptive Gaussian Quadrature (AGQ) [Pinheiro and Bates, 2002, Plan et al., 2012]. In practice, the objective is to locate the quadrature grids where the integrand values are significantly different from 0 not to waste computational resources for nodes where the integrand values collapse to almost 0. To achieve this goal, the quadrature grids are located around the maximum of the integrand. Furthermore, the variables are scaled with curvature information given by the Hessian matrix (instead of the covariance matrix), allowing to adapt the nodes to the local behavior of the integrand function. Thus, the likelihood is transformed as:

$$\begin{aligned}
\mathcal{L}_i(\boldsymbol{\Psi}) &= \int_{\mathbb{R}^d} \alpha_i(\boldsymbol{\Psi}) \exp\left(-\left(\tilde{g}_i(\mathbf{y}_i, \boldsymbol{\theta}_i, \boldsymbol{\Psi})\right)\right) \mathrm{d}\boldsymbol{\theta}_i \\
&= \int_{\mathbb{R}^d} \alpha_i(\boldsymbol{\Psi}) \exp\left(-\left(\tilde{g}_i(\mathbf{y}_i, \mathbf{b}_i, \boldsymbol{\Delta}_i)\right)\right) \mathrm{d}\mathbf{b}_i \\
&= \int_{\mathbb{R}^d} \alpha_i(\boldsymbol{\Psi})\omega_i^d |\mathbf{L}(\mathbf{y}_i, \widehat{\mathbf{b}}_i, \boldsymbol{\Psi})| \exp\left(-\left(\tilde{g}_i(\mathbf{y}_i, \widehat{\mathbf{b}}_i + \omega_i \mathbf{L}(\mathbf{y}_i, \widehat{\mathbf{b}}_i, \boldsymbol{\Psi})\mathbf{z}^*, \boldsymbol{\Psi}) \right.\right. \\
&\quad \left.\left. + \frac{\|\mathbf{z}^*\|^2}{2}\right)\right) \exp\left(-\frac{\|\mathbf{z}^*\|^2}{2}\right) \mathrm{d}\mathbf{z}^* \ .
\end{aligned} \tag{2.43}$$

In Eq. (2.43), $\mathbf{L}(\mathbf{y}_i, \widehat{\mathbf{b}}_i, \boldsymbol{\Psi})$ refers to the Cholesky decomposition of $\mathcal{H}(g_i)(\widehat{\mathbf{b}}_i)^{-1}$ and variable $\mathbf{z}^*$ is defined by $\mathbf{z}^* := (\omega_i \mathbf{L}(\mathbf{y}_i, \widehat{\mathbf{b}}_i, \boldsymbol{\Psi}))^{-1}(\mathbf{b}_i - \widehat{\mathbf{b}}_i)$. In the first line of Eq. (2.43), the change of variable corresponding to Eq. (2.10) is performed. In the last line, the change of variable that maps $\mathbf{b}_i$ into $\mathbf{z}^*$ is performed with Jacobian determinant $\omega_i^d|\mathbf{L}(\mathbf{y}_i, \widehat{\mathbf{b}}_i, \boldsymbol{\Psi})|$, followed by a multiplication and division by $\exp\left(-\frac{\|\mathbf{z}^*\|^2}{2}\right)$. Then, $\mathbf{z}^*$ is again distributed with $\mathcal{N}(\mathbf{0}_{\mathbb{R}^d}, \mathbf{I}_d)$. As for standard Gaussian quadrature, the interest of variable $\mathbf{z}^*$ is that its coordinates are independent from one another. A similar formula as Eq. (2.42) can then be applied to estimate the likelihood.

The number of nodes is defined by the user. If $N_{GQ} = 1$, it is tantamount to the Laplacian approximation [Pinheiro and Bates, 2002]. This setting is important as it defines the precision of the estimation (at the expense of significant increase of the computational cost). As the number of evaluations of function $g_i(\cdot)$ is N_{GQ}^d for each individual, [Pinheiro and Bates, 2002] suggests not to go beyond 7 for computational cost reasons. The SAS software [SAS, 2004] that proposes functions for calibration with mixed-effects models integrates a procedure to automatically tune this number. Starting from $N_{GQ} = 1$, this number is increased iteratively until reaching a convergence criterion on the estimation of the likelihood defined by the user.

Yet, this method is limited by several aspects in addition to the problems of accuracy. For instance, it is heavily sensitive to an increase in the number of parameters (though more advanced methods can be implemented [Smolyak, 1963]) and it depends on the regularity properties of integrand [Vonesh et al., 2002]. Moreover, the computational cost increases if strong dependencies between the model parameters arise. Furthermore, the adaptive version suffers from the same limitations as the Laplace approximation approach (it applies in the same conditions) but requires stronger numerical conditions to apply. Indeed, if the inverse of $\mathcal{H}(g_i)(\widehat{\mathbf{b}}_i)$ is not positive-definite, then, it has no Cholesky decomposition and the method cannot proceed. Note that also the available derivations in the literature are limited to cases where the model parameters distribution is Gaussian multivariate [Plan et al., 2012].

The methods presented in this section to approximate the likelihood call for a direct maximization, equivalently, the minimization of the opposite log-likelihood that is easier to achieve because of numerical matters. Given the possible issues, this minimization should be carried out carefully. The next section discusses the available solutions to achieve this minimization.

2.3.3 Maximizing the likelihood

The maximization of the likelihood is a central issue. The maximization problem of the likelihood provides the maximum likelihood estimate (MLE) of $\boldsymbol{\Psi}$ labeled $\widehat{\boldsymbol{\Psi}}$ that reads as:

$$\widehat{\boldsymbol{\Psi}} = \underset{\boldsymbol{\Psi}\in\Gamma_{\boldsymbol{\Psi}}}{\arg\min} - \ln(\mathcal{L}(\boldsymbol{\Psi})) \ , \tag{2.44}$$

with $\Gamma_{\boldsymbol{\Psi}}$ the admissible set of the full population parameters. In theory, to maximize the likelihood (or rather minimize the opposite log-likelihood), the usual optimization algorithms (*e.g.* gradient descent, Bayesian optimization, evolutionary algorithms, Expectation-Maximization (EM) algorithms) are available [Song et al., 2005]. A quasi-newton method is proposed in [Nocedal and Wright, 2006], applied in [Guedj et al., 2007] and implemented in [SAS, 2004]. Fisher scoring is discussed in [Longford, 1987] for multi-level mixed-models, the EM algorithm is applied in [Pinheiro et al., 2001] and the simplex method is chosen in [Aarons, 1993]. Nevertheless, remember that generally, the likelihood of mixed-effects does not have a closed form expression and requires to be estimated with one of the methods described earlier. Thus, the minimization method should be robust enough with respect to the possible errors of the estimation of the likelihood function, and depends thus on the approximation method. Within the techniques mentioned earlier, it appears that the methods involving sampling suffer from variability: two successive estimations of the likelihood with (MCMC-)IS for the same full population parameters would not give exactly the same likelihood values (that requires to set the random seed). Of course, this numerical variability decreases with the number of samples, but it requires extra computational resources (note also that the choice of the sampling density may influence the variability of the likelihood estimation) [Song et al., 2005]. Thus, performing likelihood minimization with the (MCMC-)IS approach requires a significant number of samples to achieve sufficient accuracy which does not seem compatible with the use of gradient descent that requires the cost function to have smoothness properties. In addition, the eventual auxiliary minimization of the AGQ and of the Laplace approximation brings its own variability, that still usually remains lower than what can be encountered with sampling-based techniques. Another difficulty that comes with gradient-descent lies in the estimation of the derivatives. With (A)GQ and the Laplace approximation, the derivatives of the approximation with respect to $\boldsymbol{\Psi}$ can be estimated analytically from the approximation provided either in Eqs. (2.42) or (2.27) (up to a proper estimation of the Hessian matrix of the behavior law [Almquist et al., 2015] as available in the SAS software [SAS, 2004], but which remains difficult with most behavior law). On the contrary, if the likelihood is estimated with first-order methods, the gradient should be approximated given it has no closed form and an approximation such as finite differences should be employed. Provided that the convergence of gradient descent partly depends on the accuracy of the estimation of the derivatives, a sufficient level of precision should be ensured. When no analytical expression is available, automatic differentiation routines [Kristensen et al., 2016, Nocedal and Wright, 2006] may be employed, at the expense of an increased numerical complexity. Furthermore, not to get trapped into local minima, it is often necessary to launch several minimizations with different initializations and implement, for instance, a multi-start version [Marti et al., 2010] of the gradient-descent. Not to be limited by the smoothness properties of the log-likelihood, gradient-free optimization can be preferred, using for instance evolutionary algorithms [Ashlock, 2006] or the simplex method, whose main advantage is to usually provide an

estimation of a local minimum at reasonable costs. On the contrary, the evolutionary algorithms are stochastic algorithms that aim to determine the global minimum combining exploration and exploitation sequences. Their main drawback is their computational cost as they usually need to perform many (costly) likelihood evaluations. In addition, because of their stochastic nature, it is more difficult to guarantee the repeatability of the results. Bayesian optimization [Frazier, 2018] is another possible choice for the minimization of the opposite log-likelihood. It does not depend upon prior assumptions and can process log-likelihood with altered regularity properties that is approximated with a Gaussian Process. This brings a first limit of the approach: we cumulate an error from the likelihood approximation and the Gaussian Process surrogate. Another limitation comes from its difficulty to achieve maximization in high dimensions, which is here the case. Indeed, provided that most continuous distributions are described by at least two parameters, for d model parameters, there are always at least $2d+1$ full population parameters to estimate ($2d$ for $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ and another for the residual model).

With mixed-effects, one can also find in the literature other options [Lavielle, 2014, Lindstrom and Bates, 1990] such as alternate schemes that combine successive maximization and approximation phases to estimate the full population parameters instead of directly carrying out the minimization of the log-likelihood. This is developed in the next section.

2.3.4 Estimating the model parameters distribution without direct likelihood maximization

The main difficulty with the likelihood comes from the computation of the integral that arises because the individual parameters are unobserved. Yet, if they were known, $\boldsymbol{\Psi}$ could be estimated by jointly maximizing the product of $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}), 1 \leqslant i \leqslant n$. This remark suggests to use sequential procedures that in a first step estimate individual parameters and in a second update the full population parameters. Two versions of such procedure can be found: the Lindstrom & Bates algorithm [Lindstrom and Bates, 1990] and the Stochastic Approximation Expectation Maximization (SAEM) algorithm [Kuhn and Lavielle, 2004]. In fact, instead of directly maximizing the likelihood, both algorithms update successively the full population parameters by resolving approximate maximizations of the likelihood.

2.3.4.1 Lindstrom-Bates algorithm [Lindstrom and Bates, 1990]

This algorithm enables to estimate the full population parameters for Gaussian distributed residuals and when $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ is a multivariate Gaussian distribution. It is based on successive approximations that alternate between estimates of the individual parameters $\boldsymbol{\theta}_i$ and fixed-effects $\boldsymbol{\mu}$ in a first step and of the covariance matrix $\boldsymbol{\Sigma}$ in a second time. More precisely, in the k^{th}-step, the individual parameters and the fixed-effects are determined by minimizing, for given proposals of covariance matrix and variances of the error term ω_i^2s embedded in $\boldsymbol{\Delta}_{i,k}$, the following criterion, with $\tilde{g}_i(\cdot)$ defined in Eq. (2.20) ,

$$\widehat{\boldsymbol{\mu}}_k, \widehat{\mathbf{b}}_{1,k}, \ldots, \widehat{\mathbf{b}}_{n,k} \coloneqq \underset{\Gamma_{\boldsymbol{\mu}} \times \Gamma_{\mathbf{b}}^n}{\arg\min} \sum_{i=1}^{n} \tilde{g}_i(\mathbf{y}_i, \boldsymbol{\mu} + \mathbf{b}_i, \boldsymbol{\Delta}_{i,k}) \ , \tag{2.45}$$

with $\Gamma_{\boldsymbol{\mu}}$ the set of admissible means and $\Gamma_{\mathbf{b}}$ the set of admissible deviations. Let us define pseudo-observations $\mathbf{w}_{i,k}, i \in [\![1, n]\!]$ as:

$$\mathbf{w}_{i,k} \coloneqq \mathbf{y}_i - \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{b}}_{i,k}) + \widehat{\mathbf{X}}_{i,k}\widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{Z}}_{i,k}\widehat{\mathbf{b}}_{i,k} \ , \tag{2.46}$$

with $\widehat{\mathbf{X}}_{i,k} \coloneqq \frac{\partial \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{b}}_{i,k})}{\partial \boldsymbol{\mu}}$ and $\widehat{\mathbf{Z}}_{i,k} \coloneqq \frac{\partial \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{b}}_{i,k})}{\partial \mathbf{b}_i}$. Equation Eq. (2.46) defines a linear model for $\mathbf{w}_{i,k}$ with design matrices $\widehat{\mathbf{X}}_{i,k}$ and $\widehat{\mathbf{Z}}_{i,k}$ [Pinheiro and Bates, 2002]. From Eq. (2.6b),

$\mathbf{b}_i \underset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_{\mathbb{R}^d}, \mathbf{\Sigma})$, it follows that the $\mathbf{w}_{i,k}$s are Gaussian distributed with mean $\widehat{\mathbf{X}}_{i,k}\widehat{\boldsymbol{\mu}}_k$ and covariance matrix $\mathbf{S}_i(\mathbf{\Psi})$ where $\mathbf{S}_i(\mathbf{\Psi}) = \omega_i^2\left(\mathbf{I}_{N_i} + \widehat{\mathbf{Z}}_{i,k}\mathbf{\Delta}_i^{-1}\mathbf{\Delta}_i^{-\top}\widehat{\mathbf{Z}}_{i,k}^{\top}\right)$. To estimate the Cholesky decomposition of the covariance matrix, $\mathbf{\Delta}_{i,k}$, and the noise variances, ω_i^2s, the following likelihood of linear mixed-effects is maximized:

$$\begin{aligned} -\ln(\mathcal{L}^k_{\text{LME}}(\mathbf{\Psi})) = \frac{1}{2}\sum_{i=1}^{n}\big(\ln(|2\pi\mathbf{S}_i(\mathbf{\Psi})|) + \\ +(\mathbf{w}_{i,k} - \widehat{\mathbf{X}}_{i,k}\widehat{\boldsymbol{\mu}}_k)^{\top}\mathbf{S}_i(\mathbf{\Psi})^{-1}(\mathbf{w}_{i,k} - \widehat{\mathbf{X}}_{i,k}\widehat{\boldsymbol{\mu}}_k)\big) \ , \end{aligned} \tag{2.47}$$

To avoid bias in the estimation of the covariance matrix and of the noise variances, a restricted version of the likelihood can be preferred [Lindstrom and Bates, 1990]. These steps are repeated until a certain convergence criterion is met.
Here, two minimizations have to be performed: the minimization of the sum of functions $\overline{g}_i(\cdot)$ defined in Eq. (2.45) and of the log-likelihood of linear mixed-effects defined in Eq. (2.47). The first phase is generally solved with a gradient based algorithm (for instance the Gauss-Newton algorithm in [Pinheiro, 2022]). The second phase can be carried out either with Newton-Raphson [Jennrich and Schluchter, 1986] or the EM algorithm [Laird and Ware, 1982, Lindstrom and Bates, 1990]. During the first iterations of the algorithm, the use of EM is recommended in [Fu et al., 2013] in the first steps to accelerate convergence, before being replaced by the Newton-Raphson algorithm to save computational time. As the first order techniques, this approach is limited to cases where the model parameters distribution is Gaussian, otherwise the expression of Eq. (2.47) is not available.

2.3.4.2 Stochastic Approximation Expectation Maximization (SAEM) [Kuhn and Lavielle, 2004, Lavielle, 2014]

When the maximization of the likelihood is complicated to perform, the Expectation-Maximization (EM) algorithm [Dempster et al., 1977] can be implemented, in particular in the case of missing data. This fits our situation, where the missing data refers to the individual parameters $\boldsymbol{\theta}_i$. It is composed of two phases, the estimation (E) step and the maximization (M) step that are carried out successively upon reaching convergence. Let us now describe the different steps of the EM algorithm. With samples of individual parameters, the following quantity can be evaluated at the $(k+1)^{\text{th}}$ iteration (E-step):

$$Q_{k+1}(\mathbf{\Psi}) \coloneqq \mathbb{E}\left[\ln(f(\overline{\mathbf{y}}, \overline{\boldsymbol{\theta}}|\mathbf{\Pi}, \boldsymbol{\omega}))|\overline{\mathbf{y}}, \mathbf{\Pi}_k, \omega_k\right] = \int_{\mathbb{R}^{nd}} \ln(f(\overline{\mathbf{y}}, \overline{\boldsymbol{\theta}}|\mathbf{\Pi}_k, \boldsymbol{\omega}_k))f(\overline{\boldsymbol{\theta}}|\overline{\mathbf{y}}, \mathbf{\Pi}_k, \boldsymbol{\omega}_k)\mathrm{d}\overline{\boldsymbol{\theta}} \ , \tag{2.48}$$

before being maximized to define the next iterate $\mathbf{\Psi}_{k+1}$ (M-step)

$$\mathbf{\Psi}_{k+1} = \underset{\mathbf{\Psi}\in\Gamma_{\mathbf{\Psi}}}{\arg\max}\, Q_k(\mathbf{\Psi}) \ . \tag{2.49}$$

In Eq. (2.48), $\overline{\mathbf{y}} \coloneqq [\mathbf{y}_1, \ldots, \mathbf{y}_n]$ is the concatenation of all output data, $\overline{\boldsymbol{\theta}} = [\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_n]$ the concatenation of all individual parameters, $\boldsymbol{\omega} \coloneqq [\omega_1, \cdots, \omega_n]$, and $\Gamma_{\mathbf{\Psi}}$ the set of admissible full population parameters representing mathematical and physical constraints. The sequence of $(\mathbf{\Psi}_k)_{k\in\mathbb{N}^*}$ converges to the MLE estimator of $\mathbf{\Psi}$ under general requirements [Wu, 1983], namely the continuity of $Q_k(\mathbf{\Psi})$ with respect to $\mathbf{\Psi}$ and $\mathbf{\Psi}_k$.
Yet, for non-linear models, $Q_k(\cdot)$ cannot be computed in closed form. Consequently, it is necessary to approximate it. With samples from $f(\overline{\boldsymbol{\theta}}|\overline{\mathbf{y}}, \mathbf{\Pi}_k, \boldsymbol{\omega}_k)$, $Q_{k+1}(\cdot)$ can estimated by Monte-Carlo integration (E-step). Given that $f(\overline{\boldsymbol{\theta}}|\overline{\mathbf{y}}, \mathbf{\Pi}_k, \boldsymbol{\omega}_k)$ is expressed as the product of $f(\boldsymbol{\theta}_i|\mathbf{y}_i, \mathbf{\Pi}_k, \omega_{ik})$ for independent individuals, the generation of samples $f(\boldsymbol{\theta}_i|\mathbf{y}_i, \mathbf{\Pi}_k, \omega_{ik})$ can be achieved as exposed in Section 2.3.2.3 describing the MCMC-IS technique. This is the Monte-Carlo Expectation Maximization (MCEM) algorithm presented in [Wei and Tanner, 1990].

The SAEM algorithm [Delyon et al., 1999] replaces the E-step of the MCMC algorithm by a stochastic approximation. Given N_{IS} samples of $\tilde{\boldsymbol{\theta}}$ drawn from $f(\overline{\boldsymbol{\theta}}|\overline{\mathbf{y}}, \boldsymbol{\Pi}_k, \boldsymbol{\omega}_k)$ denoted $\left\{\tilde{\boldsymbol{\theta}}_{k,\ell}\right\}_{\ell\in[\![1,N_{IS}]\!]}$, it approximates $Q_{k+1}(\cdot)$ as

$$Q_{k+1}(\boldsymbol{\Psi}) \coloneqq Q_k(\boldsymbol{\Psi}) + \tau_k(\ln(\prod_{\ell=1}^{N_{IS}} f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\tilde{\boldsymbol{\theta}}_{k,\ell}|\boldsymbol{\Pi})) - Q_k(\boldsymbol{\Psi})) \ , \tag{2.50}$$

which is then maximized as in Eq. (2.49) to get the next iterate $\boldsymbol{\Psi}_{k+1}$. In Eq. (2.50), $(\tau_k)_{k\in\mathbb{N}^*}$ is a sequence of decreasing weights that must comply with some requirements to ensure the convergence of the sequence $(\boldsymbol{\Psi}_k)_{k\in\mathbb{N}}$ to the MLE $\widehat{\boldsymbol{\Psi}}$ along with others regularity conditions on the likelihood [Kuhn and Lavielle, 2004], giving $\lim_{k\to+\infty} \boldsymbol{\Psi}_k = \widehat{\boldsymbol{\Psi}}$. The main interest of the procedure is that for distributions in the exponential family, the approximation simplifies into a Robbins-Monro scheme [Herbert Robbins and Monro, 1951]. Furthermore, for a Gaussian distribution, the resolution of the maximization problem defined in Eq. (2.49) is analytic. Its main drawback lies in the sampling of $f(\overline{\boldsymbol{\theta}}|\overline{\mathbf{y}}, \boldsymbol{\Pi}_k, \boldsymbol{\omega}_k)$ with costly MCMC methods at each step of the algorithm.

2.3.5 A brief comparison of the calibration routines

One of the main issues for techniques that involve MCMC sapmpling are monitoring of the convergence and the computational cost. These already are difficult issues that are made even harder to handle with mixed-effects than with standard applications. Indeed, remember that such methods are employed at each step of the algorithms in which they are involved and a relevant tuning should fit all iterations. Furthermore, it is necessary to strike a balance between the precision of the integrals implied in the criterion formulation (that depends on the number of samples) and the available computational facilities. Apart from their own issues, MCMC-IS and SAEM techniques also inherit the advantages and drawbacks of Monte-Carlo based approaches. At the expense of a significant computational cost, the only requirement is that $f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i)$ is integrable with respect to $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})$. Otherwise, in a general formulation, MCMC-IS and SAEM can be applied even if the previous densities have poor regularity properties or a complicated numerical behavior. Note that handling non-smooth quantities may require more samples to get a proper estimation of the integrals and a more complex control of the convergence of the MCMC methods. However, the interest of SAEM decreases if the model parameters distribution does not belong to the exponential family as the approximation and maximization phases become significantly harder to carry out. Finally, none of these two approaches demand any approximation of the integrand (*e.g.*, linearization) that would require further assumptions.

Quadrature techniques are appealing as they do not require to make any approximation of the integrand. They are slightly less general than Monte-Carlo based techniques as they require stronger regularity assumptions. Like MCMC-IS techniques, they require large number of model evaluations (N_{GQ}^d for each individual, to be added to those necessary for the resolution of the minimization defined in Eq. (2.30) for each individual if the adaptive version is chosen). In addition, even if the adaptive formulation provides a more accurate evaluation of the likelihood function by focusing the integrand evaluations on the region centered on the estimated individual parameters, it is limited by the same requirements encountered in the Laplace approach and the IS approach suggested by [Pinheiro and Bates, 1994]. Among the drawbacks lie their heavy sensitivity to the curse of dimensionality, *i.e.* the accuracy of the estimation at a fixed number of grid points decreases as the number of model parameters increases. More advanced techniques (*e.g.* using sparse grid [Smolyak, 1963]) may help to solve this issue but do not appear to be widespread for the estimation of the likelihood of mixed-effects.

The first order and Lindstrom-Bates methods are the easiest to implement. Aside to their potential lack of precision, they are limited by the fact that they can only be implemented if $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ is a Gaussian distribution. This is a strong limitation because in practical cases, the exact distribution remains unknown and is not necessarily Gaussian.

The Laplace approximation of the likelihood also suffers from several weaknesses. For instance, it relies on restrictive assumptions that are difficult to guarantee from a numerical point of view (*e.g.* positive-definiteness of the Hessian matrix of $g_i(\cdot)$ at $\widehat{\mathbf{b}}_i$). Remember also that with few observations available (small N_i), the validity of the approximation can be challenged.

Except for first order methods and crude Gaussian quadrature, all methods that approximate or maximize the likelihood require to perform at least an auxiliary optimization (Laplace approximation, Lindstrom-Bates algorithm, IS from [Pinheiro and Bates, 1994] and AGQ) or an MCMC sampling (SAEM and MCMC-IS from [Lavielle, 2014]). Given that the maximization of the likelihood relies on these auxiliaries methods, they should be employed cautiously as they may flaw the estimation of the full population parameters. This implies other issues about the convergence and the accuracy of the results from these auxiliary methods that are difficult to assess throughout the optimization of the global likelihood.

Numerically, comparative reviews of the methods have been carried out for the calibration of pharmacokinetic models. In [Pinheiro and Bates, 1994], IS, the Laplace approximation, the quadrature techniques in its standard and adaptive version and Lindstrom-Bates have been compared. 5 full model parameters have to be calibrated. As it could be expected, the IS approximation of the likelihood proposed in [Pinheiro and Bates, 1994] requires too much model evaluations to be considered as efficient. For comparable results, it demanded about 10^7 model evaluations while calibration with the Laplace approximation only required about 10^4. Crude Gaussian quadrature showed its limits. For instance, one estimate of standard deviation was 10% above those provided by other techniques with 10 nodes (in each direction). This gap was filled by increasing the number of nodes to 100 that made the number of model evaluations soar to 10^7. On the contrary, AGQ only demanded 3×10^4 model evaluations to reach convergence, underlining the interest of centering the integration interval on the estimates of individual parameters. The fastest method was the Lindstrom-Bates algorithm with only 2×10^3 model evaluations to achieve similar results as the other methods. Others case study in [Pinheiro and Bates, 1996] seemed to show that the Laplace approximation and AGQ provide more accurate estimations of the full population parameters than the Lindstrom-Bates algorithm. Besides, [Pinheiro and Bates, 1994] recommends to use the Lindstrom-Bates procedure to get a first estimation of $\boldsymbol{\Psi}$ that is next refined with a calibration with the Laplace approximation.
In another study conducted on a similar type of models [Plan et al., 2012], mixed-effects calibration for pharmacokinetic was studied with the Laplace approximation, SAEM and AGQ methods with different settings of the algorithms, (*e.g.* number of iterations, detection of convergence, initial guess, *etc.*), different model errors and "rich" and "sparse" data ("rich" data refers to data with 4 observation points and "sparse" data refers to data with 2 observation points randomly selected from the "rich" data). This study concluded that both SAEM and Laplace provided good results at a reasonable costs. AGQ was the slowest method : for 100 simulated data-sets, the median of model evaluations reached 6.74×10^{11} to be compared to a median of 1.17×10^{11} model evaluations for the different SAEM setings and a median of 10^{10} model evaluations for the Laplace approach over the 9 scenarios, limiting this approach to "simple models". However, AGQ provided the best results in terms of accuracy: for all scenarios, the ratio of the root mean squared relative errors divided by the best one was below 1.5 contrary to other techniques that sometimes provided results with more errors. Furthermore, the SAEM was found to be dependent on the initial estimate of the full population parameters. In particular,

it [Plan et al., 2012] showed that switching from exact to altered initial conditions dropped the completion rate [1] by 30% in average for one of the implementation of the SAEM algorithm. On the contrary, the Laplace approach minimized with a gradient algorithm had a completion rate of 100%, whichever situation was considered.
Though it is not as general as the MCMC-IS approach, the Laplace approach is fast in spite of the auxiliary optimization and provides results with acceptable accuracy for different phenomenological model [Pinheiro and Bates, 1994]. This will be the chosen method in the rest of this work.

Several routines devoted to the mixed-effects calibration of phenomenological model are available : nlmixr [Fidler et al., 2019], Monolix [Lavielle, 2008], nmlefit in Matlab [nml, 2008], SAS [SAS, 2004], with a more complete list available in [Stegmann et al., 2007]. Table 2.1 summarizes the different techniques discussed along with their advantages and drawbacks.

Table 2.1: Comparison of the different techniques for the estimation of the full population parameters.

	Advantages	Drawbacks
AGQ	Provides accurate estimation of the full population parameters	Difficult to apply for non-Gaussian $f_{\boldsymbol{\Theta}\|\boldsymbol{\Pi}}$. Subjected to the curse of dimensionality
(MCMC-)IS	Most accurate and general method	Most computationally intensive method
First-order methods	Fastest estimation of the full population parameters. Easy to implement	Raw approximation of the likelihood. Limited to Gaussian $f_{\boldsymbol{\Theta}\|\boldsymbol{\Pi}}$
Laplace approximation	Good compromise between accuracy of the results and computational costs	Applies for a restricted range of model parameters distributions
Lindstrom-Bates	Fast estimation of the full population parameters with acceptable accuracy	Limited to Gaussian $f_{\boldsymbol{\Theta}\|\boldsymbol{\Pi}}$
SAEM algorithm	Applies for most type of assumptions on $f_{\boldsymbol{\Theta}\|\boldsymbol{\Pi}}$ in theory	Computationally intensive. In practice, the difficulty of the implementation increases much for non-Gaussian $f_{\boldsymbol{\Theta}\|\boldsymbol{\Pi}}$

The methods to estimate the likelihood have been described. They make it possible to minimize the opposite log-likelihood of mixed-effects to estimate both the population and individual parameters. Once the full population parameters are estimated, then the individual parameters (of practical interest in engineering problems) can be inferred, as described next.

2.3.6 Estimating the individual parameters

The determination of the individual parameters is a key issue to ensure a proper likelihood maximization and should not be neglected. To explain its importance, it should be underlined that mixed-effects simply aims to tune the population parameters ($\boldsymbol{\Pi}$) in order to have the best agreement possible between the assumed model parameters distribution ($f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$) and the empirical distribution of the individual parameters. Thus, the data should provide enough information to estimate accurately the individual parameters. In particular, the model should be

[1] Completion is defined as the proportion of achievements of the calibration over the 100 data-sets (*i.e.* the minimization was not interrupted by numerical issues) and does not account for the quality of the results.

appropriately chosen and not present imperfections (*e.g.* not over-parameterized). Otherwise, if the individual parameters cannot be learnt with sufficient precision, the regularization term ($f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})$ in Eq. (2.15) or $\|\boldsymbol{\Delta}\mathbf{b}_i\|^2$ in function $g_i(\cdot)$ in Eq. (2.18)) will coerce them with the mean value and flaw the estimation of the standard-deviations and correlations parameters. In fact, generally, it appears that to ensure a proper estimation of a given marginal (the univariate distributions of each coordinate), the corresponding parameter should be properly estimated over the different individuals, and the same occurs for the parameters involved in coupling relations.
Note that all issues encountered within the classical identification techniques described in Chapter 1 remain with mixed-effects. In particular, problems such as model identifiability, over-fitting or uncertainty on the individual parameters should be studied. Those identifiability concerns are of prime importance and can be illustrated for the Laplace approximation or IS without MCMC sampling. Indeed, those ones involve the Hessian of function $g_i(\cdot)$, and with the chosen assumptions, it comprises the matrix $\frac{\partial \mathcal{F}(\mathbf{t}_i,\widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i}^\top \frac{\partial \mathcal{F}(\mathbf{t}_i,\widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i}$ which is in fact the sensitivity matrix. Thus, if the model is non-identifiable, both the Hessian matrix and the regularization term will flaw the estimation of the population parameters. However, it should be noted that sensitivity analyses can help to solve such issues by screening prior to calibration the non-identifiable parameters.
From a practical point of view, the individual parameters can be either estimated with MCMC sampling or with minimization. With MCMC sampling, individual parameters are predicted as the mean (or the mode) of the PDF $\pi_i(\boldsymbol{\theta}_i|\boldsymbol{\Psi})$, *i.e.*, the empirical mean of $f(\boldsymbol{\theta}_i|\mathbf{y}_i,\boldsymbol{\Pi},\omega_i)$ (the maximum of $f(\boldsymbol{\theta}_i|\mathbf{y}_i,\boldsymbol{\Pi},\omega_i)$) [Lavielle, 2014]. MCMC samples can also be propagated through the model. Individual parameters can also be computed through the minimization of function $g_i(\cdot)$ at $\widehat{\boldsymbol{\Psi}}$ as described in Eq. (2.30). Once $\boldsymbol{\Psi}$ is chosen, the individual parameters $\boldsymbol{\theta}_i$ can be estimated by looking for the dominating contribution in the individual likelihood $\mathcal{L}_i$ (Eq. (2.21)) through the minimization of function $g_i(\cdot)$ (Eq. (2.18))

$$\widehat{\boldsymbol{\theta}}_i = \underset{\mathbf{b_i}\in\Gamma_{\boldsymbol{\theta}}}{\arg\min}\, g_i(\mathbf{y}_i,\boldsymbol{\theta}_i,\boldsymbol{\Psi}) \ , \tag{2.51}$$

with $\Gamma_{\boldsymbol{\theta}}$ the search space of the individual deviations.

The population and individual parameters can be estimated with the methods we have described. We can now discuss the analysis of the results.

2.4 Studying the results of mixed-effects calibration

Once the maximization of the likelihood is achieved, calibration results should be validated. The first aspect that should be studied concerns the estimation of the full population parameters. Aside from matters of bias and error on the estimation of full population parameters that can be only estimated with synthetic data, the choice of the model including the choice of the distribution of the random-effects can be investigated. Furthermore, given that the number of available individuals is finite, the full population parameters are flawed with uncertainty. Its quantification *via* either the asymptotic theory or bootstrap can be represented by confidence regions on $\boldsymbol{\Psi}$. In addition, given that the estimation of the full population parameters is made to best fit the empirical distribution of the individual parameters, the calibration of the individuals should also be checked. This demands assessing the residual model and also checking that the data allows to properly estimate the different model parameters to prevent, in particular, overfitting.

2.4.1 Studying the population level

. The objective of mixed-effects is to provide an estimation of the model parameters $\boldsymbol{\Psi}$ for the distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$. With synthetic data, when the associated hyperparameters $\boldsymbol{\Psi}$ are known, results are analyzed with usual indicators that correspond to bias, mean squared error [Pinheiro and Bates, 1994] and Kullback-Leibler (KL) divergence [Kullback and Leibler, 1951]. The advantage of the latter quantity comes from the fact that it is a joint criterion that seems more relevant with respect to the simultaneous estimation of the distribution parameters. However, it is difficult to exploit this quantity for anything other than a comparison between two divergences. In particular, it is difficult to interpret them from a physical point of view.
Given the impossibility to compute the exact likelihood (except for linear models, Gaussian distributed residuals and Gaussian distributed model parameters) and the finite number of specimens, the estimation of the full population parameters is uncertain. This uncertainty can be estimated with the same tools as those presented in Chapter 1, that is to say asymptotic theory and bootstrap in the frequentist framework. The major issue for both of these approaches is that they are based on an asymptotic behavior reached for a large number of individuals. Given the costs of experiments, few individuals are usually available in material characterization problems and, most of the time, the asymptotic behavior is not met. This issue is examined in [Thai et al., 2013]. The asymptotic theory is based on the behavior of the MLE estimator that is asymptotically Gaussian with mean the MLE estimate and covariance matrix the inverse of the Fisher Information Matrix (FIM):

$$\boldsymbol{\Psi} \underset{n\to+\infty}{\sim} \mathcal{N}\left(\widehat{\boldsymbol{\Psi}}, \frac{1}{n}\mathcal{I}_n(\widehat{\boldsymbol{\Psi}})^{-1}\right) , \tag{2.52}$$

with $\mathcal{I}_n(\widehat{\boldsymbol{\Psi}}) \coloneqq -\mathbb{E}_{\boldsymbol{\Psi}}\left[\frac{\partial^2 \ln(\mathcal{L}(\widehat{\boldsymbol{\Psi}}))}{\partial\boldsymbol{\Psi}\partial\boldsymbol{\Psi}^\top}\right]$. Another approximation of the distribution of the MLE estimator (not employed in the mixed-effects methods literature) can be implemented [Barndorff-Nielsen, 1983]. Both approximations require the estimation of the FIM. Provided that the computation of the likelihood is tricky, its estimation should be performed cautiously. As the exact FIM is expressed as an expectation, it can be estimated with integration methods based on linearizations [Mentré et al., 1997], Laplace approximation [Vong et al., 2014] or adaptive Gaussian quadrature (AGQ) [Nguyen and Mentré, 2014]. In [Riviere et al., 2016], a procedure is proposed allowing to estimate the FIM with great accuracy, at the expense of almost prohibitive computational costs (it mixes Monte-Carlo and MCMC sampling at two different levels). Confidence regions can also be deduced from likelihood ratio techniques [Stram and Lee, 1994]. A simpler approach consist in approximating the exact FIM with the observed FIM. The latter can be estimated with finite differences [Lavielle, 2014] or with the help of the Louis formula [Louis, 1982] for procedures based on the EM algorithm.

Bootstrap also applies to estimate uncertainties on the full population parameters [Das and Krishen, 1999, Thai et al., 2013, Thai et al., 2014]. The same convergence theorems as those for the individual approach of Chapter 1 apply about the empirical distribution of the bootstrap samples. It can be performed in its parametric and non-parametric versions. The general principle remains the same as proposed in the first chapter (see Section 1.4.2 in Chapter 1), only the sampling procedure changes. The difference with the bootstrap for standard calibration is that re-sampling is conducted at both the individual and population levels (*i.e.* between the observations of each individual for the first level and between the individuals for the second). It is employed for instance in [Higgins et al., 1997, Jolling et al., 2005, Wu and Zhang, 2002]. Two types of bootstrap are usually chosen for mixed-effects models: non-parametric (sometimes called case bootstrap) and parametric bootstrap [Das and Krishen, 1999, Thai et al., 2013, Thai et al., 2014]. As in its standard version described in Chapter 1, parametric bootstrap aims to create new data based on the residual model (here Gaussian distributed with homoscedastic

variance) and on the model parameters distribution. To create the bootstrapped data at step k, (that consists here on new sets of individuals $\{\mathbf{y}_{\text{boot},i,k}\}_{i\in[\![1,n]\!]}$), model parameters $\{\boldsymbol{\theta}_{i,k}\}_{i\in[\![1,n]\!]}$ are generated from the calibrated distribution $f_{\boldsymbol{\Theta}|\widehat{\boldsymbol{\Pi}}}$. Then, the exact outputs $\{\mathcal{F}(\mathbf{t}_i,\boldsymbol{\theta}_{i,k})\}_{i\in[\![1,n]\!]}$ are computed, to which are added noise profiles $\{\boldsymbol{\xi}_{i,k}\}_{i\in[\![1,n]\!]}$ sampled from the distribution of the residuals (here $\mathcal{N}(\mathbf{0}_{\mathbb{R}^{N_i}},\widehat{\omega_i}^2\mathbf{I}_{N_i})$).

Bootstrap for mixed-effects can also be performed in a non-parametric fashion [Das and Krishen, 1999, Thai et al., 2013, Thai et al., 2014]. Following a first calibration (including the estimation of the individual parameters $[\widehat{\boldsymbol{\theta}}_i]_{i\in[\![1,n]\!]}$ and of the residuals $[\widehat{\mathbf{r}}_i]_{i\in[\![1,n]\!]}$), the individual parameters are then resampled as $\{\widehat{\boldsymbol{\theta}}_{\text{boot},i,k}\}_{i\in[\![1,n]\!]}$ as $\widehat{\boldsymbol{\theta}}_{\text{boot},i,k}\coloneqq\widehat{\boldsymbol{\theta}}_{U_{i,k}}, i\in[\![1,n]\!]$ with $\{U_{i,k}\}_{i\in[\![1,n]\!]}$ uniform numbers defined as $U_{i,k}\sim\mathcal{U}([\![1,n]\!])$. The bootstrapped individuals account for the sum of the models for $\{\widehat{\boldsymbol{\theta}}_{\text{boot},i,k}\}_{i\in[\![1,n]\!]}$ (in fact $\{\mathcal{F}(\mathbf{t}_i,\widehat{\boldsymbol{\theta}}_{\text{boot},i,k})\}_{i\in[\![1,n]\!]}$) to which are added the bootstrapped residuals $\mathbf{r}_{i,\text{boot},k}$, themselves resampled independently as described in Section 1.4.2 of Chapter 1 for each individual. To construct confidence intervals, two solutions can be chosen [Carpenter and Bithell, 2000]. The first consists in ordering the bootstrap samples. The $(1-\alpha)100\%$ confidence interval is defined by the $\alpha 100^{\text{th}}\%$ and $(1-\alpha)100^{\text{th}}\%$ quantiles of the empirical bootstrap distribution. The quantiles are given by the $(k+1)\alpha^{\text{th}}$ and $(k+1)(1-\alpha)^{\text{th}}$ elements of the ordered bootstrap samples. Another possibility is to approximate the empirical distribution of the bootstrap samples with a Gaussian distribution giving, for the confidence interval $I^l_{1-\alpha}$ at level $1-\alpha$ of each component of vector $\boldsymbol{\Psi}$ labeled $[\Psi_l]_{l\in[\![1,p]\!]}$

$$I^l_{1-\alpha}\coloneqq[\Psi_{l,\text{med}}-q_{1-\frac{\alpha}{2}}\text{sd}(\Psi_{l,\text{boot}}),\Psi_{l,\text{med}}+q_{1-\frac{\alpha}{2}}\text{sd}(\Psi_{l,\text{boot}})]\ , \tag{2.53}$$

with $\boldsymbol{\Psi}_{\text{med}}$ the median of the bootstrap samples, $q_{1-\frac{\alpha}{2}}$ the quantile associated to $1-\alpha$ and $\text{sd}(\Psi_{l,\text{boot}})$ the empirical standard deviation of the bootstrap samples defined as:

$$\text{sd}(\Psi_{l,\text{boot}})\coloneqq\sqrt{\frac{1}{N_B}\sum_{k=1}^{N_B}(\Psi_{l,\text{med}}-\Psi_{l,\text{boot},k})^2}\ , \tag{2.54}$$

with N_B the number of bootstrap iterations.

The mixed-effects framework also provides tools for model testing. With mixed-effects, the "model" may refer to the residual model, to the behavior law $\mathcal{F}$ or to the model of the individual parameter, that is to say to their distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$. To be more precise, the latter refers to the choice of marginals and of the dependence structure. It can be achieved by comparing Bayesian Information Criteria (BICs) [Delattre et al., 2015] or Akaike Information Criteria (AICs) [Greven and Kneib, 2010, Vaida and Blanchard, 2005] given the same behavior law. Note that the expression of AICs and BICs is changed from the ones that apply for standard calibration, notably the expression of the penalty term depending on whether model parameters are fixed or random. Cross-validation also remains available [Colby and Bair, 2013] for hypotheses testing with mixed-effects. Still, these indicators are derived under an asymptotic behavior (in terms of number of individuals) and should be used carefully if such an assumption is not met. Mixed-effects calibration is strongly impacted by the choice of the model parameters distribution that determines the choice of the method to approximate the likelihood, the maximization algorithm and the individual parameters. Statistical tests can be carried out to challenge the chosen hypotheses. For instance, it may not be legitimate to consider all of the model parameters as random. This question is investigated in [Stram and Lee, 1994] in which a test is proposed to check whether parameters should be held fixed or not for linear mixed-effects.

More generally, the choice of the distribution of the random-effects can be challenged. To validate it, [Verbeke and Molenberghs, 2013] proposes a statistical test. It is based on a "gradient"

function $\mathfrak{G}(\cdot,\cdot)$ that is computed as follows:

$$\mathfrak{G}(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}},\mathbf{b}) \coloneqq \frac{1}{n}\sum_{i=1}^{n}\frac{f(\mathbf{y}_i|\mathbf{b},\boldsymbol{\Pi},\omega_i)}{\mathcal{L}_i(\boldsymbol{\Psi})} \ . \tag{2.55}$$

In Eq. (2.55), $\mathbf{b}$ refers to the candidate deviations. Contrary to previous equations, there is not index i attached to it in order to show that contrary to the $\mathbf{b}_i$s, it is not sampled from $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$. This quantity can be read as the average of likelihood ratios. Each ratio seeks to determine whether it is more probable to observe individual i on its own (numerator of Eq. (2.55)), or if the specimen is more likely drawn from a population of specimens (denominator of Eq. (2.55)) in which the variability of the mechanical properties is modeled by $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$. Thus, if the assumption on $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ is true, this function complies with $\mathfrak{G}(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}},\mathbf{b}) \leqslant 1 \ \forall \ \mathbf{b} \in \mathbb{R}^d$ [Verbeke and Molenberghs, 2013]. From this observation, [Drikvandi, 2017] proposes a test based on the statistic T evaluated with Monte-Carlo based techniques given by:

$$T = \int_{\mathbb{R}^d} (\widehat{\mathfrak{G}}(f_{\boldsymbol{\Theta}|\widehat{\boldsymbol{\Pi}}},\mathbf{b}) - 1)^2 f_{\boldsymbol{\Theta}|\widehat{\boldsymbol{\Pi}}}(\mathbf{b}|\widehat{\boldsymbol{\Pi}})\mathrm{d}\mathbf{b} \ . \tag{2.56}$$

In Eq. (2.56), $\widehat{\mathfrak{G}}$ refers to the empirical estimation of $\mathfrak{G}$ that is computed after the achievement of the calibration of the population parameters $\widehat{\boldsymbol{\Psi}}$. Asymptotically, under H_0 (the assumption on $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ is correct), T follows a weighted sum of independent chi-squared distributions with one degree of freedom each. Yet, with limited number of individuals, parametric bootstrap may be preferred to sample its distribution to invalidate or not H_0 [Drikvandi, 2017].

2.4.2 Studying the individual level

In addition to the full population parameters, mixed-effects enable to estimate the individual parameters. With synthetic data, when the exact individual parameters $\boldsymbol{\theta}_i$s are known, the results are analyzed with usual indicators corresponding to bias and mean squared error. These quantities can be estimated either between the parameters (in the parameters space) or between the exact/simulated data and the model for the calibrated individual parameters (in model space).
Contrary to the model parameters distribution whose nature (shape, dependence structure, *etc.*) is unknown in practical cases, it is possible to validate the estimation of the individual parameters to some extent. For instance, the model for the calibrated individual parameters can be confronted to the experimental data. Note that given the presence of measurement errors, the identification of model parameters is uncertain. Once the full population parameters are calibrated, samples from density $f(\boldsymbol{\theta}_i|\mathbf{y}_i,\widehat{\boldsymbol{\Pi}},\widehat{\omega})$ can be drawn with MCMC sampling as proposed in [Lavielle, 2014] to estimate the impact of measurements errors on the individual parameters, enabling to determine trust regions and plotting confidence intervals. The residual model can be also checked: distribution, independence, *etc.* These assumptions can be checked with the same tools as described in Chapter 1.

Confidence in the individual parameters may also be evaluated with shrinkage indicators [Savic and Karlsson, 2009]. Among them can be found the η-shrinkage, defined as $\mathrm{sh}(\eta) \coloneqq 1 - \frac{\widehat{\mathbb{V}}(\widehat{\theta}_{il})}{\mathbb{V}(\theta_{il})}$, with $\widehat{\mathbb{V}}(\widehat{\theta}_{il})$ the empirical variance of the estimated individual parameters and $\mathbb{V}(\theta_{il})$ the estimated variance by maximum likelihood of the model parameters $\theta_l, l \in [\![1,d]\!]$ (it is in vector $\boldsymbol{\Psi}$). The empirical variance of the estimated $\widehat{\theta}_{il}$ can be either obtained directly from the calibrated individual parameters or determined following a more complex procedure proposed in [Lavielle, 2014] which is discussed below.

If the data is informative enough, the empirical variance of the identified individual parameter will be close to its variance estimated by ML, and $\mathrm{sh}(\eta)$ would be close to zero. On the

contrary, sh(η) different from zero indicates that data does not allow for a proper identification. This may occur when the model output is not sensitive enough to the parameter in question and the value of the individual parameters is driven by $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})$ in the likelihood (see Eq. (2.15)) which can be seen as a regularization term. Indeed, this term coerces the individual parameters $\boldsymbol{\theta}_i$ to be as close as possible to the modes/means of the model parameter distribution. In such case, the θ_{il}s are concentrated around the means. Note that sometimes the ratio of variances is replaced by the ratio of the standard deviations. An unbiased estimation of the η-shrinkage is presented in [Lavielle, 2014] for the estimation of $\widehat{\mathbb{V}}(\theta_{il})$. This estimation is performed in two steps. Given the calibrated population parameters, the first step consists in generating individuals $\tilde{\mathbf{y}}_i$ following Eqs. (2.6a) and (2.6b). Then, from these individuals, the distribution $f(\boldsymbol{\theta}_i|\boldsymbol{\Pi}, \boldsymbol{\omega}, \tilde{\mathbf{y}}_i)$ is sampled by MCMC as described in the section dedicated to the MCMC-IS, enabling to estimate $\widehat{\mathbb{V}}(\theta_{il})$ from the pooled samples. Even if this procedure provides accurate and unbiased estimates of the η-shrinkage, it involves successive MCMC samplings, with each of them to be checked before exploiting the generated samples. Thus, it makes this method complicated to set up, in addition to the fact that the computational burden of the procedure can be prohibitive for complex behavior laws.

2.5 Conclusion

This chapter was devoted to the presentation of population approaches for the calibration of behavior laws under material variability. The statistical modeling presented explicitly accounts for individual variability thus changing the way to perform its estimation. Among these formulations, two main approaches may be distinguished: Hierarchical Bayesian Model and mixed-effects. Because of its Bayesian nature, HBMs require most of the time nested MCMC samplings which usually become intractable for complex behavior laws. This explains why mixed-effects methods are chosen in our work. For this purpose, the derivation of the likelihood for such mixed-effects models has been presented considering different assumptions, along with the appropriate estimation techniques. A focus has been made on the Laplace approximation approach that provides an approximation of the likelihood with reasonable accuracy and limited computational costs (and which will be thus the selected approximation method). The minimization methods for the approximated opposite log-likelihood has been described, along with the methods for the auxiliary minimization that arises in the implementation of the likelihood computation with the Laplace approach. Finally, the ways to analyze the results both at the population and individual level have also been presented. Mixed-effects methods appear to be a promising framework for the estimation of material variability as it both provides results at the population and individual levels. Once the results are examined with the approaches presented in this chapter, the estimated variability can be propagated for analysis of practical interest, for example to size structures. It is important to remember that mixed-effects are a generalization of the standard approaches. Indeed, if the variance of the model parameters shrinks to 0, the individual parameters are equal to their mean almost surely.
The current implementations of population approaches are mainly limited to simple behavior laws. To be more precise, the model of interest is usually analytical with few parameters involved (up to 4 in many pharmacokinetics models, see for instance [Lavielle, 2014, Mentré et al., 1997, Pinheiro and Bates, 1994]). In structural applications, there can be more parameters to estimate (8 in [Papadimas and Dodwell, 2021], 7 in [Economides et al., 2021], 5 in [Song et al., 2020]) but this remains much smaller than the 26 parameters of the Onera Progressive Failure Model (OPFM) [Laurin, 2005], that adds up to the multi-axial nature of the composite behavior. Another major issue specific to the mechanical model is that they describe different behaviors: elasticity, viscosity, damage, *etc.* to which are associated subsets of parameters determined on specific experiments. In fact, instead of considering the behavior law as a black-box function and calibrating the vector of parameters as a whole, a more relevant and efficient way to carry out

the model identification is to consider separated blocks of model parameters dedicated to each physical phenomenon. A sequential scheme already implemented in the individual approach (that is to say without consideration of the specimen variability) [Jung et al., 2022, Youn et al., 2011] seems more appropriate to solve calibration problem, but requires an adaptation to be applied in population approaches, especially for the management of the correlation coefficients. This refers to our final target in order to pave the way to the identification of complex non-linear models such as the OPFM model.
The next two chapters are devoted to the presentation of this thesis contributions. The third chapter consists in a verification of the employability of the mixed-effects approach in the context of the calibration of material behavior laws. An emphasis is put on the impact of the number of available specimens on the estimation of the population parameters. In the last chapter, an extension of the all-at-once to the sequential calibration method is proposed and illustrated on a test-case, allowing to predict the distribution of a design criterion: the failure stress of an open hole plate loaded in tension.

CHAPTER 3

CHARACTERIZING MATERIAL VARIABILITY USING MIXED-EFFECTS

Contents

Chapter contributions:

- Implementation of the mixed-effects methodology for uniaxial nonlinear mechanical behavior laws
- Illustration of the reduction of the uncertainty of the population parameter estimates with the number of individuals
- Investigation of the necessity to account for statistical dependencies between parameters despite having few individuals
- Testing the calibration of non-Gaussian models of the parameters distribution
- Illustration of the ability of mixed-effects methodology to characterize material variability from 13 replications of experimental tensile tests achieved on Ceramic Matrix Composite material

3.1 Introduction

The study of material variability is a long-standing issue for composite materials [Defense et al., 1999]. In the first chapter, its impact has been highlighted, even if the usual methods [Gallagher and Doherty, 2007, Kennedy and O'Hagan, 2001] do not apply to estimate it. In the second chapter, a literature review of existing population-based approaches [Demidenko, 2013, Lavielle, 2014] and more specifically mixed-effects methods [Pinheiro and Bates, 2002] has been done. These methods have been widely implemented in many fields and in particular, in pharmacokinetics to determine the intrinsic variability that arises through the responses to drug injections on various subjects [Lavielle, 2014]. Yet, most of the phenomenological models used in these fields do not present specific complexities (no implicit formulation, *etc.*) or are linearized, making the opposite log-likelihood much easier to minimize, contrary to those that are implemented in mechanics, such as the Onera Damage Model (ODM) [Maire et al., 1997, Marcin et al., 2011, Ben Ramdane, 2014]. Consequently, it is necessary to verify the applicability of the mixed-effects statistical framework presented in the previous chapter for the calibration of behavior laws such as the ODM model with repetitions of specimens.
One objective of this chapter is to study the influence of settings and hypotheses on mixed-effects techniques for the calibration of behavior laws with synthetic data. In particular, a focus is made on the number of available individuals, which is a key feature in the design of experimental campaigns in terms of financial costs and time considerations. Another objective is to apply mixed-effects techniques on real strain-stress curves. Indeed, the effective goal is to characterize

the variability from existing materials. In addition, experimental data presents challenges (presence of model bias, experimental issues, *etc.*) that do not arise with synthetic data. This chapter begins with a brief recap of mixed-effects calibration among which a remainder of the statistical assumptions and of the likelihood computations, followed by a quick presentation of the ODM behavior law in Section 3.2. In the Section 3.3, the test protocol for mixed-effects is presented, including the generation of synthetic data performed by first considering a multivariate Gaussian distribution as the exact probability distribution of the model parameters. Section 3.4.1 proposes a first inquiry of the calibration with the mixed-effects of the ODM model with the generated synthetic data, knowing the parameter distributions and the parameter realizations associated to the synthetic data. Once the approach has been validated on this fully controlled case, the choice of the algorithm used to carry out the auxiliary optimization is discussed in Section 3.4.2 in order to reduce the computational costs without downgrading the estimation of the model parameters distribution. The influence of the number of individuals on the estimation of the population parameters and the associated uncertainty is studied in Section 3.4.3. The inclusion and estimation of a dependency structure within the multivariate Gaussian distribution characterizing material variability is investigated in Section 3.4.4. However, in practice, the nature of the distribution of the model parameters is unknown. In Section 3.4.5, the same virtual data are treated with mixed-effects models for different assumptions on the marginals. Finally, the thirteen real strain-stress curves of a Ceramic Matrix Composite material presented in the first chapter are processed with mixed-effects with different modeling assumptions (on the shape of the variability and on the residual model) in Section 3.5 to assess the applicability of the method to real data.

3.2 Definition of the calibration problem with mixed-effects

This section consists of a brief recap of the definition of the calibration problem, probabilistic modeling (Section 3.2.1) and of the ODM behavior law (Section 3.2.2).

3.2.1 Statistical framework of the calibration problem with mixed-effects

As stated in Chapter 2, the population approaches aim to characterize the variability that arises through repetitions of experiments and demand to model both the population and the individual levels. Let's concentrate on the case of scalar measures such as those used in Chapter 3. Note n the number of available repetitions with i the corresponding index, N_i the number the observations of the i^{th} subject with j the index of the j^{th} observation. The set of model parameters of dimension d (there are d parameters to calibrate) that characterizes the i^{th} specimen is labeled $\boldsymbol{\theta}_i$. The j^{th} input and output data of the i^{th} specimen are respectively labeled t_{ij} and y_{ij}. The behavior law is denoted $\mathcal{F}(\cdot,\cdot)$. The noise term is labeled ξ_{ij}. As discussed in the previous part, the probability distribution that represents material variability, called the model parameters distribution, is labeled $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$. The equations governing the mixed-effects are [Demidenko, 2013, Lavielle, 2014, Pinheiro and Bates, 2002]:

$$\begin{cases} \forall i \in [\![1,n]\!], \boldsymbol{\theta}_i \underset{\text{i.i.d.}}{\sim} f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}} \ , & \text{(3.1a)} \\ \forall (i,j) \in [\![1,n]\!] \times [\![1,N_i]\!] \ , y_{ij} = \mathcal{F}(t_{ij},\boldsymbol{\theta}_i) + \xi_{ij} \ , & \text{(3.1b)} \\ \forall (i,j) \in [\![1,n]\!] \times [\![1,N_i]\!] \ , \xi_{ij} \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\omega_{ij}^2) \ , & \text{(3.1c)} \end{cases}$$

with ω_{ij} the standard-deviation of the noise of the j^{th} output measure of the i^{th} individual and $\boldsymbol{\Pi}$ stands for to the population parameters (the hyperparameters of the model parameters probability distribution) with s its dimension. Within the mixed-effects framework, the individual parameters are decomposed between fixed-effects $\boldsymbol{\beta}$ and random-effects $\mathbf{b}_i$ as:

$$\boldsymbol{\theta}_i = \boldsymbol{\beta} + \mathbf{b}_i \ . \tag{3.2}$$

To limit the number of parameters to estimate, the noise is supposed to be homoscedastic, *i.e.* $\omega_{ij} = \omega_i \; \forall \; j \in [\![1, N_i]\!]$ [Pinheiro and Bates, 2002]. Finally, the vector of parameters to be calibrated is denoted $\boldsymbol{\Psi}$:

$$\boldsymbol{\Psi} \coloneqq \begin{bmatrix} \boldsymbol{\Pi}, & \omega_1, & \ldots, & \omega_n \end{bmatrix} \; .$$

The dimension of $\boldsymbol{\Psi}$ is denoted p ($p = n + s$). The population-based approaches seek $\boldsymbol{\Psi}$ and provide an estimate of the individual parameters $\{\boldsymbol{\theta}_i\}_{i \in [\![1,n]\!]}$ as a by-product. The full population parameters are estimated by maximizing a likelihood of $\boldsymbol{\Psi}$, that is to say $\mathcal{L}(\boldsymbol{\Psi}) = f(\mathbf{y}_1, \cdots, \mathbf{y}_n | \boldsymbol{\Psi})$, with f a generic letter for probability density functions (PDFs). The likelihood reads as:

$$\mathcal{L}(\boldsymbol{\Psi}) = \prod_{i=1}^{n} \int_{\mathbb{R}^d} f(\mathbf{y}_i | \boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i | \boldsymbol{\Pi}) \mathrm{d}\boldsymbol{\theta}_i \; . \tag{3.3}$$

As detailed in the previous chapter, Section 2.3.2.2, the opposite log-likelihood is approximated with the Laplace approximation as follows:

$$\begin{aligned} -\ln(\mathcal{L}(\boldsymbol{\Psi})) &= \sum_{i=1}^{n} \ln \left(\int_{\mathbb{R}^d} f(\mathbf{y}_i | \boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i | \boldsymbol{\Pi}) \mathrm{d}\boldsymbol{\theta}_i \right) \\ &\approx \sum_{i=1}^{n} \left(\frac{1}{2} \ln(|\mathcal{H}(g_i)(\widehat{\boldsymbol{\theta}}_i)|) + g_i(\mathbf{y}_i, \widehat{\boldsymbol{\theta}}_i, \boldsymbol{\Psi}) \right) - \frac{d}{2} \ln(2\pi) \; , \end{aligned} \tag{3.4}$$

with function $g_i(\cdot)$ defined as:

$$g_i(\mathbf{y}_i, \boldsymbol{\theta}_i, \boldsymbol{\Psi}) \coloneqq -\ln(f(\mathbf{y}_i | \boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i | \boldsymbol{\Pi})) \; . \tag{3.5}$$

In Eq. (3.4), $\widehat{\boldsymbol{\theta}}_i$ refers to the solution of the following minimization problem:

$$\widehat{\boldsymbol{\theta}}_i = \underset{\boldsymbol{\theta}_i \in \Gamma_{\boldsymbol{\theta}}}{\arg\min} \, g_i(\mathbf{y}_i, \boldsymbol{\theta}_i, \boldsymbol{\Psi}) \; , \tag{3.6}$$

with $\Gamma_{\boldsymbol{\theta}}$ the set of admissible parameters representing physical and mathematical constraints. Finally, in Eq. (3.4), $\mathcal{H}(g_i)(\cdot)$ is the Hessian of the function $g_i(\cdot)$ defined in Eq. (3.5).

3.2.2 The ODM behavior law

The behavior law of interest in this chapter is a simplified uniaxial version of the ODM model [Maire et al., 1997, Marcin et al., 2011, Ben Ramdane, 2014], the same as the one implemented in the first chapter. Monotonic tension experiments are considered in the following. The input data is the strain (without units) and the output is the stress (in MPa). There are four parameters to estimate: the Young's modulus (E_0 in GPa) that stands for the elastic linear slope, the damage threshold (y_{0s} in kPa) which indicates the start of damage, the damage evolution celerity (y_{cs} in MPa) related to the kinetics of damage and the damage saturation (d_c without units) which indicates the highest possible damage. The final objective is to estimate the variability that arises from the 13 experiments of monotonic loadings carried out on CERASEP A400 material [Naslain, 2004, Viricelle et al., 2001], a composite material with ceramic matrix. These repetitions, already introduced in Chapter 1, are reproduced in Figure 3.1.

3.3 Testing protocol for mixed-effects calibration

We seek now to demonstrate the ability of mixed-effects models to estimate material variability. To illustrate this point, mixed-effects are applied first on synthetic data, allowing to estimate error criteria between exact and calibrated values. This section is organized as follows. First, data generation is exposed. Then, the maximization of the likelihood and the assessment methodology used for comparison are detailed.

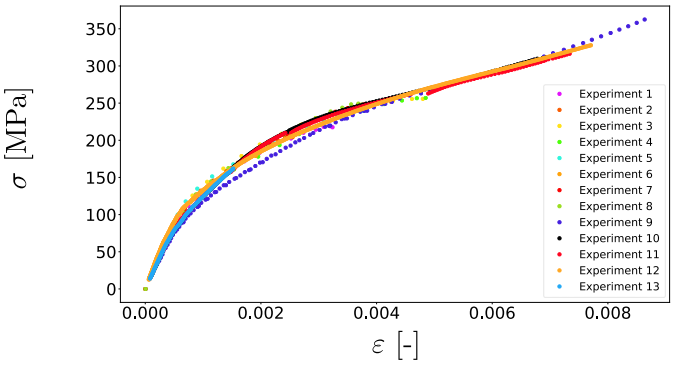


Figure 3.1: 13 experimental monotonic tensile strain-stress curve on a Ceramic Matrix Composite material [Naslain, 2004, Viricelle et al., 2001].

3.3.1 Generating synthetic data

A multivariate distribution, defined by population parameters $\mathbf{\Pi}$, is chosen to generate synthetic data. This set of parameters will be denoted $\mathbf{\Pi}_{\text{exact}}$ in the following. Individual parameters $\boldsymbol{\theta}_{i,\text{exact}}$ are sampled from this distribution to compute, for each of these specimens, the ODM model outputs $\boldsymbol{\sigma}_{i,\text{exact}}$. A noise profile is added to this data to obtain the synthetic data involved in the calibration process $\boldsymbol{\sigma}_{i,\text{noisy}}$. The mixed-effects approach is then applied to the synthetic data and the results obtained are confronted to exact values.
In this work, the model parameters distribution is chosen as a multivariate Gaussian distribution, a usual modeling [Bright et al., 2011b, Niyigena et al., 2016], as given fixed average and second orders, this distribution has the highest entropy [Cover and Thomas, 2006]. In other words, it is the most general one. In addition, with such an assumption, the estimation of the likelihood is also easier. Such a distribution is parameterized by both a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\mathbf{\Sigma}$. For the sake of simplicity, the model parameters are first considered independent (hence $\mathbf{\Sigma}$ is a diagonal matrix). Table 3.1 presents the values of the latter which define the multivariate Gaussian distribution used to generate the synthetic data (before the application of noise). The values of the population parameters refer to usual values for this type of material. In particular, the mean values refer to the average of the parameter values independently calibrated on each repetition. The standard-deviations are chosen to enable the expression of material variability

Table 3.1: Exact values of the parameters distribution corresponding to $f_{\mathbf{\Theta}|\mathbf{\Pi}_{\text{exact}}}$.

	E_0 [GPa]	y_{0s} [kPa]	y_{cs} [MPa]	d_c [-]
Mean	189	25.0	2.24	3.77
Standard-deviation	17.9	8.00	0.439	0.893

The virtual data are generated by sampling model parameters from the target distribution $f_{\mathbf{\Theta}|\mathbf{\Pi}_{\text{exact}}}$. The model outputs $\{\boldsymbol{\sigma}_{i,\text{exact}}\}_{i\in[\![1,n]\!]}$ are computed for each individual. The input strain profile starts from $\varepsilon = 0$ up to $\varepsilon = 0.008$, the maximum strain reached by the ninth experiment in Figure 3.1, regularly sampled with 29 points (slightly more points than most of the experiments depicted in Figure 3.1 to ensure the applicability of the Laplace approximation). All the specimen have the same number of observations so that all of them have the same weight within the likelihood function ($N_i = 29\ \forall i \in [\![1,n]\!]$). An heteroscedastic noise is added to the experimental data:

$$\boldsymbol{\sigma}_{i,\text{noisy}} = \boldsymbol{\sigma}_{i,\text{exact}} \times (1 + \tau\boldsymbol{\xi}) \text{ with } \boldsymbol{\xi} \sim \mathcal{N}(0, \mathbf{I}_{N_i}) \text{ and } \tau = 0.02\ , \tag{3.7}$$

with $\mathbf{I}_{N_i}$ the identity matrix of $\mathcal{M}_{N_i}(\mathbb{R})$. An heteroscedastic noise is preferred to its homoscedastic counterpart. Indeed, with an homoscedastic noise, in the linear part which corresponds to

the lowest level of stress, the contribution of the noise to the resulting observations could be higher than the model output, preventing a proper identification of Young's modulus in particular. Exact and noisy data are depicted in Figure 3.2.

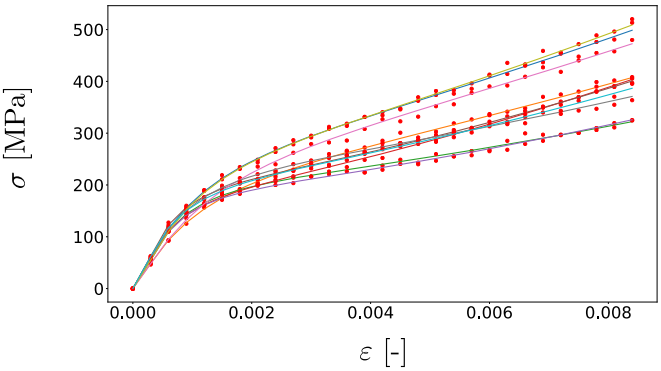


Figure 3.2: Example set of 10 synthetic stress-strain curves, exact models (lines) and noisy data (dots).

In the calibration process, the noise is considered as homoscedastic across, both, time and specimens, because considering heteroscedasticity would require to design an appropriate noise model and exhibit sufficient statistics to justify it. In addition, this allows to see whether the homoscedastic assumption is sufficient to provide relevant estimations of the population parameters or whether the heteroscedastic hypothesis cannot be simplified. Formally, this can be written as:

$$\forall (i,j) \in [\![1,n]\!] \times [\![1,N_i]\!] \ \omega_{ij}^2 = \omega^2 \ .$$

Consequently, the matrix $\mathbf{\Omega}$ is reduced here to ω and $\mathbf{\Psi}$ to $\mathbf{\Psi} = [\mathbf{\Pi}, \omega]$.

3.3.2 Numerical implementation

This section presents the numerical settings for the implementation of the mixed-effects methods. The implementation of the code has been carried out in Python with the NumPy library [Harris et al., 2020] used for matrix calculation and the coding of the behavior law and with the OpenTURNS library [Baudin et al., 2016] for the probabilistic modeling. The implementation of the minimization of the global likelihood approximated in Eq. (3.4) and the estimation of the individual parameters that has to be carried out for each likelihood estimation are discussed next. The pseudocode is detailed in the appendices A.

3.3.2.1 Minimization of the global likelihood

To estimate the population parameters and the residual model all assembled within $\mathbf{\Psi}$ (the full population parameters), it is necessary to maximize the likelihood of mixed-effects models or rather to minimize the opposite of the log-likelihood, which can be translated into the following optimization problem:

$$\widehat{\mathbf{\Psi}} = \underset{\mathbf{\Psi} \in \Gamma_{\mathbf{\Psi}}}{\arg\min} - \ln(\mathcal{L}(\mathbf{\Psi})) \ , \tag{3.8}$$

with $\Gamma_{\mathbf{\Psi}}$ the set of admissible full population model parameters that consists here into search bounds. The maximum likelihood estimate (MLE) of $\mathbf{\Psi}$ denoted $\widehat{\mathbf{\Psi}}$ is decomposed between the population estimates $\widehat{\mathbf{\Pi}}$ and the calibrated residual model $\widehat{\omega}$. These search bounds are reported in Table 3.2. If these search bounds are hit, they can be updated if deemed necessary. The search space on the population parameters is set up so that the mid-point corresponds almost to the exact value of the population parameter. The search bounds also intend to permit an exploration of the parameter space. Those on ω are have been determined so that they contain the variance of the synthetic noise. In the remainder of this manuscript, the standard-deviations will be noted $\mathrm{sd}(\cdot)$, rather than using the usual letter σ that will be reserved for the stress.

Table 3.2: Search space for the full population parameters.

μ_{E_0} [GPa]	$\mu_{y_{0s}}$ [kPa]	$\mu_{y_{cs}}$ [MPa]	μ_{d_c} [-]
$[120, 235]$	$[5.00, 50.0]$	$[1.10, 3.50]$	$[1.30, 6.60]$

sd(E_0) [GPa]	sd(y_{0s}) [kPa]	sd(y_{cs}) [MPa]	sd(d_c) [-]	ω [MPa]
$[5.00, 38.0]$	$[0.500, 15.0]$	$[0.1, 2.25]$	$[0.1, 2.25]$	$[2.25,10.0]$

To carry out the minimization of the opposite log-likelihood, one has to choose the optimization algorithm. In Chapter 2, several methods were presented to perform the likelihood minimization among which one could find both gradient descent [Nocedal and Wright, 2006] and derivative free algorithms such as the evolutionary algorithms [Ashlock, 2006]. The main issue that prevents to use gradient descent lies in the difficulty in estimating accurately the derivative of the objective function. Indeed, such algorithms require to compute analytically or numerically at least the gradient of the opposite log-likelihood estimated with the Laplace approximation. To estimate these quantities, it is necessary to compute the Hessian matrix of the behavior law with respect to the model parameters [Almquist et al., 2015], that is to say $\frac{\partial^2 \mathcal{F}}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_i^\top}(\mathbf{t}_i, \cdot)$. This estimation with the simplified uniaxial ODM model is already a complicated task (and almost impossible with the full model). Thus, the only remaining option is to estimate the gradient numerically. The problem is that the estimation methods (*e.g.*, finite differences) provide noisy estimates of the opposite log-likelihood gradient and are difficult to tune effectively (in particular the choice of the differentiation step). Automatic differentiation techniques may solve this issue but require many model evaluations, with no absolute guarantee to get a proper result. Finally, remember that the objective function is not the exact likelihood but an estimate (in fact its Laplace approximation [Pinheiro and Bates, 2002]) and differentiating an approximation may therefore increase the numerical noise. Thus, we decide to carry out the log-likelihood minimization with gradient-free optimization algorithms, and in particular evolutionary algorithms [Ashlock, 2006]. Though they require many likelihood estimations, they exhibit a greater ability to reach the maximum of the likelihood, $\widehat{\boldsymbol{\Psi}}$, despite the possible presence of local minima.

The algorithm chosen here is an evolutionary algorithm called the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) algorithm [Hansen et al., 2003] because it has proven its capability to solve nonsmooth optimization problems [Hansen and Ostermeier, 2001, Hansen et al., 1997]. The optimization variables, $\boldsymbol{\Psi}$, are normalized between 0 and 1. A maximum of 750 iterations of the algorithm are allowed to maximize the likelihood. The size of the population of the CMA-ES algorithm is $\lceil 4 + 3 \times \ln(p) \rceil$ points, p being the number of full population parameters to calibrate ($\lceil \cdot \rceil$ the ceiling function) as suggested in [Hansen et al., 2003]. Given that most continuous distributions are parameterized by at least 2 parameters, there is at least $2d$ parameters to calibrate, to which must be added at least one variance of the error term if it is considered as a Gaussian white noise with the same variance for all individuals and parameters, and thus $p \geqslant 2d + 1$. Of course, including a dependency structure increases the number of population parameters to estimate. The starting point corresponds to $0.65 \times \mathbf{1}_{\mathbb{R}^p}$. Default settings for the convergence of the minimization of the CMA-ES algorithm are considered. To be more precise, if either the change in objective function or normalized candidate population parameters is smaller than 10^{-6}, the minimization stops. Other elements about CMA-ES are detailed in `http://www.cmap.polytechnique.fr/~nikolaus.hansen/html-pythoncma/cma-pysrc.html#CMAOptions.evalall`.

3.3.2.2 Estimation of the individual parameters

Remember that in practice, mixed-effects calibration for phenomenological model aims to tune the population parameters to make the model parameters distribution fit the best with the

empirical distribution of the individual parameters. Consequently, the chosen method should be able to identify properly the individual parameters on the different samples. For the Laplace approximation, this is performed through the minimization of function $g_i(\cdot)$ defined in Eq. (3.5). During the optimization of the global likelihood, it is difficult to assess the quality of the identification of the individual parameters. To ensure a proper resolution of this minimization problem, the identification of the individual parameters obtained through the solving of an auxiliary optimization problem Eq. (3.6) is achieved using the CMA-ES algorithm, with at most 500 iterations. The same rule as for the population parameters is used to determine the CMA-ES population size which is set to $\lceil 4+3\times\ln(d)\rceil$ as $\boldsymbol{\theta}_i$ is of dimension d. Let the fixed-effects refer to the most probable values of the individual parameters, i.e., the maxima of the density associated to $\boldsymbol{\Pi}$. The search space is centered around the fixed-effects. The fixed-effects also act as the initial point. The individual parameters search interval is defined as $\Gamma_{\boldsymbol{\theta}} = [0.3\boldsymbol{\beta}, 1.7\boldsymbol{\beta}]$, with $\boldsymbol{\beta}$ the fixed-effects (which reduce to the mean for a Gaussian multivariate distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$). The optimization variables are between 0 and 1.

3.3.3 Testing the mixed-effects approach with synthetic data

Once the minimization of the negative log-likelihood is achieved, both the population and individual estimates should be confronted to their exact counterpart to assess the quality of the results.
To compare the results at the population level, the relative error and the Kullback-Leibler (KL) divergence can be used as described in Chapter 2. The relative error $\mathcal{E}(\widehat{\boldsymbol{\Pi}})$ in % stands for:

$$\mathcal{E}(\widehat{\boldsymbol{\Pi}}) \coloneqq \left[\mathcal{E}(\widehat{\Pi}_1), \quad \ldots, \quad \mathcal{E}(\widehat{\Pi}_s)\right] \text{ with } \mathcal{E}(\widehat{\Pi}_l) \coloneqq 100 \times \frac{|\Pi_{\text{exact},l} - \widehat{\Pi}_l|}{\Pi_{\text{exact},l}}, \forall l \in [\![1, s]\!] \ , \tag{3.9}$$

noting s the number of population parameters. The KL divergence [Kullback and Leibler, 1951], $\text{KL}(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}_{\text{exact}}}, f_{\boldsymbol{\Theta}|\widehat{\boldsymbol{\Pi}}})$, between the exact distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}_{\text{exact}}}$ and the approximated calibrated distribution $f_{\boldsymbol{\Theta}|\widehat{\boldsymbol{\Pi}}}$ reads as:

$$\text{KL}(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}_{\text{exact}}}, f_{\boldsymbol{\Theta}|\widehat{\boldsymbol{\Pi}}}) \coloneqq \int_{\mathbb{R}^d} f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}_{\text{exact}}}(\boldsymbol{\theta}) \ln\left(\frac{f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}_{\text{exact}}}(\boldsymbol{\theta})}{f_{\boldsymbol{\Theta}|\widehat{\boldsymbol{\Pi}}}(\boldsymbol{\theta})}\right) \mathrm{d}\boldsymbol{\theta} \ . \tag{3.10}$$

This criterion is interesting because it is a joint criterion, that is to say all aspects (mean, standard-deviations and possibly correlations) are taken into account, contrary to the error indicators which focus independently on each population parameter. Having a joint quality metric is consistent with the joint calibration of the population parameters. In general, the KL divergence is computed with crude Monte-Carlo. Yet, if both the exact and calibrated model parameters distribution are Gaussian multivariate distributions of dimension d parameterized by $\boldsymbol{\Pi} = [\boldsymbol{\mu}, \boldsymbol{\Sigma}]$ and $\widehat{\boldsymbol{\Pi}} = \left[\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}\right]$ respectively, the KL divergence has a closed-form and refers to [Pardo, 2018]:

$$\text{KL}(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}_{\text{exact}}}, f_{\boldsymbol{\Theta}|\widehat{\boldsymbol{\Pi}}}) = \frac{1}{2}\left(\text{tr}(\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}) + (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}})^\top \widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) + \ln\left(\frac{|\widehat{\boldsymbol{\Sigma}}|}{|\boldsymbol{\Sigma}|}\right) - d\right) \ . \tag{3.11}$$

The adequation between the calibrated and the synthetic data can also be evaluated for each individual. To verify that the different specimens are correctly identified either in terms of calibrated individual parameters $\widehat{\boldsymbol{\theta}}_i$ or estimated outputs $\mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)$, the averaged relative error on the individual parameters $\mathbf{e}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_n)$ in %:

$$\begin{aligned} &\mathbf{e}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_n) \coloneqq \frac{1}{n}\sum_{i=1}^{n} \mathbf{e}_i(\widehat{\boldsymbol{\theta}}_i) \text{ with } \mathbf{e}_i(\widehat{\boldsymbol{\theta}}_i) = \left[e_{i1}(\widehat{\theta}_{i1}), \quad \ldots, \quad e_{id}(\widehat{\theta}_{id})\right] \\ &e_{il}(\widehat{\theta}_{il}) \coloneqq 100 \times \frac{|\theta_{il,\text{exact}} - \widehat{\theta}_{il}|}{\theta_{il,\text{exact}}} \ \forall l \in [\![1, d]\!] \ , \end{aligned} \tag{3.12}$$

and the averaged error in between the exact and estimated output $D(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_n)$:

$$D(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_n) \coloneqq \frac{1}{n} \sum_{i=1}^{n} \frac{1}{N_i} \|\boldsymbol{\sigma}_{i,exact} - \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)\|_2^2 \ , \tag{3.13}$$

are computed. D corresponds to the average of the squared differences between the model output and the observation points, and can be interpreted as an average mean squared error between the exact and calibrated curves over all the individuals. This is a usual criterion used in many calibration problems that allows to determine the accuracy of the calibration in the model space with a comparison with respect to the output data.
Furthermore, to account for the specificity of the sampled specimens, the whole calibration process is repeated 20 times with different exact individuals $\boldsymbol{\theta}_{\text{exact}}$ considering the same population parameters $\boldsymbol{\Pi}_{\text{exact}}$. The above indicators are averaged over the repetitions and their dispersion is characterized by the coefficient of variation $\text{COV}(X)$ defined for any non-zero quantity X as $\text{COV}(X) = \frac{\sqrt{\widehat{V}(X)}}{\widehat{M}(X)}$ with $\widehat{M}(X)$ the empirical mean of X and $\widehat{V}(X)$ its empirical variance.
All indicators used for the general analyses have been presented so far. It is possible now to start studying the calibration results coming from the first implementation. All results are shown with 3 significant digits. All figures displayed in this chapter correspond to a parameter distribution associated to the median of the KL divergence.

3.4 Calibrating with synthetic data

This section is dedicated to the study of mixed-effects calibration and checks that this approach can estimate properly the material variability and the individual parameters. Different aspects of mixed-effects are studied such as the influence of the number of individuals involved in the calibration process, the computational costs implied by the method, the inclusion of dependency structure between the model parameters, and the choice of non-Gaussian distribution for the marginals. The ability of mixed-effects to provide consistent estimates of material variability for different assumptions on the model parameters distribution is investigated.

3.4.1 First calibration with mixed-effects models

In this section, the synthetic data are constructed such that there is no dependency between the 4 parameters of the model and that they follow a Gaussian distribution (as in theory). The noise is considered as a Gaussian white noise with homoscedastic unknown variance. In this case, $p = 9$ parameters are calibrated: the 4 means and the 4 standard-deviations of the parameters and the variance of the noise term. The mixed-effects approach is applied to the calibration of the simplified ODM model with a number of individuals (n=10) that corresponds to the order of magnitude of repetitions usually available in mechanics [Defense et al., 1999]. The objective is to carry out a first calibration of the ODM model with mixed-effects to analyze the consequences of the choice of such a methodology. This investigation is organized as follows. The first step consists in checking that the individual parameters are accurately calibrated which is of prime importance as their estimation has a strong impact on the fit of the population parameters. The convergence of the global optimization of the likelihood, solution of the minimization problem defined in Eq. (3.8) is then detailed. Once these two steps are completed, it becomes possible to examine the calibration either at the population level (marginals) or at the specimen level (calibration of the specimens in both model space and parameter space). The calibration of the residual models ends this investigation. This process is illustrated in the flowchart in Figure 3.3.

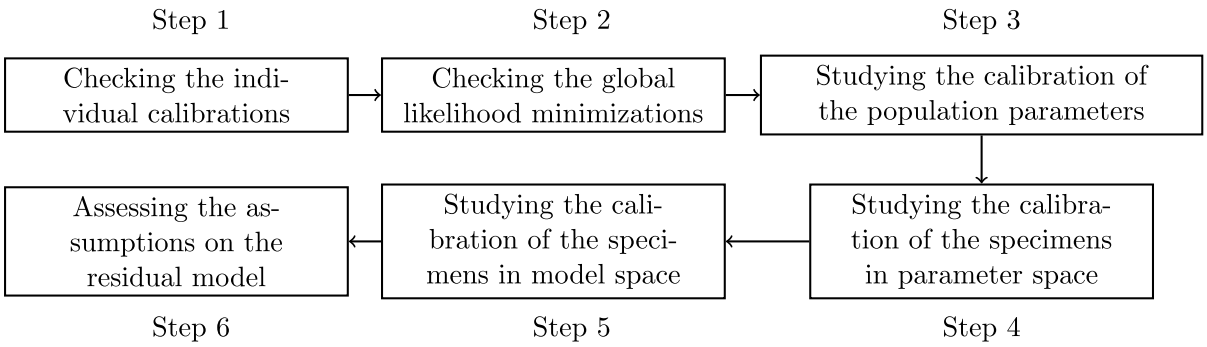


Figure 3.3: Flowchart of the different steps for the assessment of the mixed-effects calibration.

3.4.1.1 Assessing the estimation of the individual parameters

The behavior of the minimization with the CMA-ES algorithm of the fitting term within function $g_i(\cdot)$, $-\ln(f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega)$, has been studied in the first chapter. Given that function $g_i(\cdot)$ corresponds to $-\ln(f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega)$ up to a quadratic regularization term with good smoothness properties (in the case of a Gaussian multivariate distribution), we do not investigate in detail the individual minimizations. Still, it remains necessary to ensure that the CMA-ES algorithm estimates the individual parameters accurately as their estimation is at the core of the Laplace approximation. To achieve this, 1,000 samples are uniformly drawn within the bounds defined in Table 3.2. For each of these samples, the individual parameters are estimated and the relative error is computed from Eq. (3.12).

Table 3.3: For the 1st repetition and 1,000 proposals of full population parameters $\boldsymbol{\Psi}$ sampled between the bounds defined in 3.2, average relative errors $\mathbf{e}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_{10})$ in % on the individual parameters estimated with the CMA-ES algorithm.

	E_0	y_{0s}	y_{cs}	d_c
Averaged relative error on the individual parameters using the CMA-ES algorithm [%]	10.6	55.6	20.7	13.9

Table 3.3 shows that the CMA-ES algorithm finds reasonable approximations of the individual parameters except the individual thresholds $y_{0s,i}$. This mainly comes from the lack of sensitivity of the model output with respect to the damage threshold. To illustrate this point, it is possible to investigate the impact of a 10% variation of the damage threshold on the model response (so belonging to $[0.9 \times \mu_{y_{0s}}, 1.1 \times \mu_{y_{0s}}]$) and to compare it with the impact of the same variation of other parameters, for instance the damage kinetics. This is illustrated in Figure 3.4.

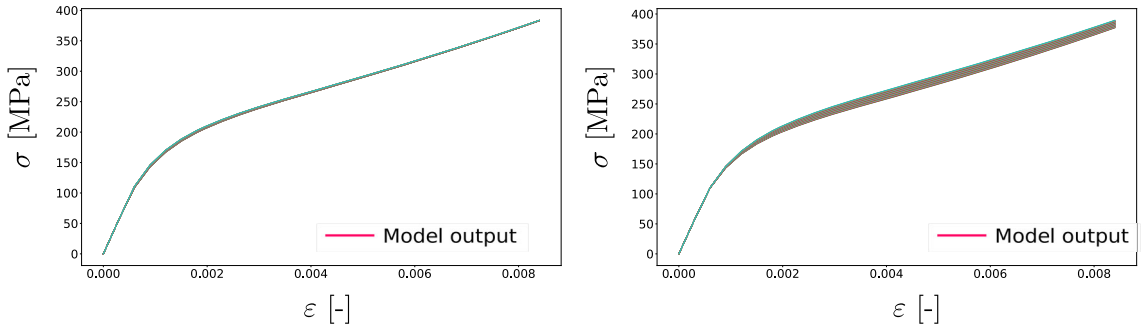


Figure 3.4: On the left, model outputs for different damage thresholds varying between $[0.9 \times \mu_{y_{0s}}, 1.1 \times \mu_{y_{0s}}]$ ($\mu_{y_{0s}}$ and other parameters correspond to the mean parameters presented in Table 3.1). On the right, model outputs for different damage kinetics varying between $[0.9 \times \mu_{y_{cs}}, 1.1 \times \mu_{y_{cs}}]$ ($\mu_{y_{cs}}$ and other parameters correspond to the mean parameters presented in Table 3.1).

Figure 3.4 confirms that the model response is less sensitive to the damage threshold than the damage kinetics, and the same occurs for the other parameters. Indeed, for the same range of variation, the model output varies of at most 5 MPa (which is small with respect to hundreds of MPa) against tens of MPa for the damage kinetics. This is consistent with experimental observations in Figure 3.1 where the material of interest exhibits a smooth transition between the linear and nonlinear regimes, making the calibration of the damage threshold difficult to perform. The estimation errors for the other model parameters remain lower than 20%. Though this can seem significant, it is necessary to remember that this identification has been carried out for proposal of population parameters that are not relevant (since the population parameters are drawn from a uniform law). In particular, this is true when the means are not appropriate and the standard-deviations small. Indeed, in such cases, the regularization prevents the auxiliary minimization to get a proper estimate of the individual parameters. Nevertheless, the likelihood for such points is small and they are not selected as possible MLE estimate. Table 3.3 also shows that the damage kinetics seems to be flawed with higher errors than for E_0 and d_c, hinting for a poorer calibration of the corresponding population parameters. This can be understood as both E_0 and d_c are easier to calibrate compared to y_{cs} as they can be almost estimated without minimization process contrary to y_{cs}. To summarize, though significant error is committed on the calibration of the damage thresholds, the estimation of the other parameters does not seem affected and thus does not prevent the estimation of the population parameters implemented next.

3.4.1.2 Minimization of the opposite log-likelihood

Before investigating the calibration results, the minimization of the likelihood should be studied to check whether the optimization process has been carried out without issues as the opposite log-likelihood minimization is challenging to perform due to the approximation of the objective function. This point can be examined by looking at the history of the opposite log-likelihood values throughout the minimization, as depicted in Figure 3.5.

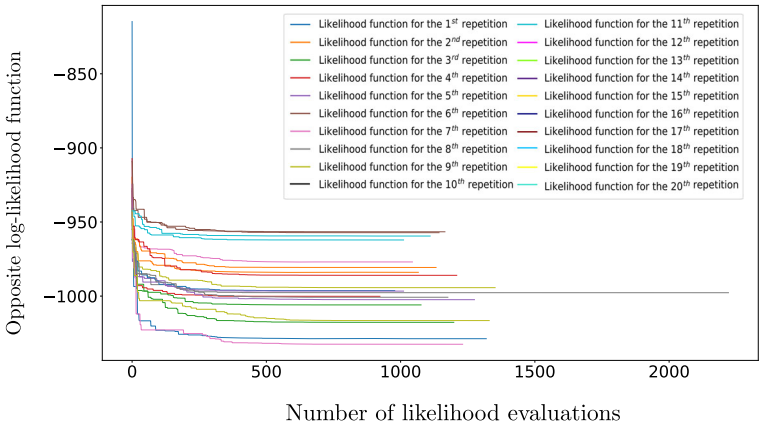


Figure 3.5: For all repetitions, best likelihood values at each evaluation. A likelihood evaluation refer to a likelihood calculation.

Figure 3.5 shows that, individually, for all repetitions, the minimization converges to an MLE estimate. Note that it is different for each repetition as the MLE depends on the data, different for each repetition, explaining why the final values of the optimizations are different. Still, the range of variation of the converged opposite log-likelihoods is about 10% of the best one and can be considered as limited. Notice that after 400 iterations, the likelihood does not evolve much compared to the previous iterations. The important remark is that in this case, no specific issue arises during the minimization process for all repetitions, demonstrating the proper convergence of the minimizations. It can also be studied at the level of the population parameters. This is achieved by recording the parameter values explored by the optimization algorithm. These (normalized) candidate parameters are displayed in Figure 3.6 for the 1$^{\text{st}}$ repetition.

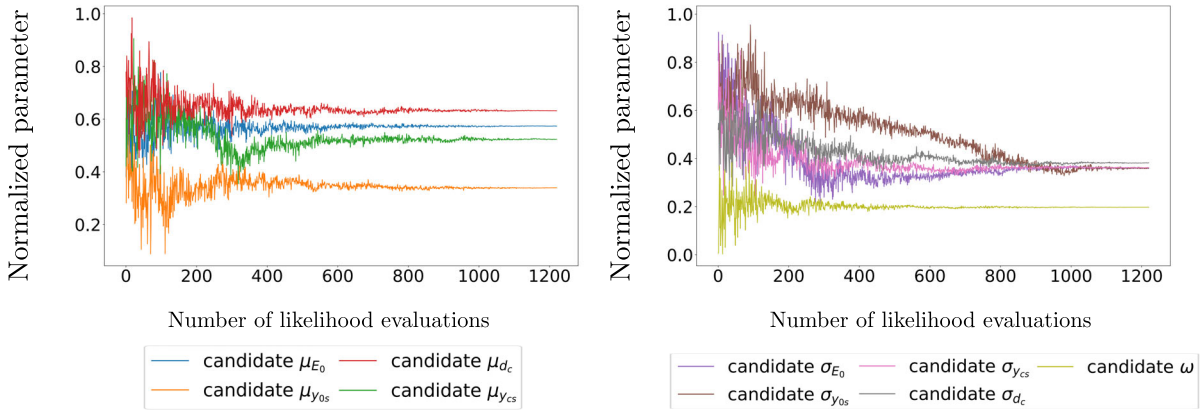


Figure 3.6: Candidate normalized parameters through the optimization for the 1st repetition. On the left, convergence of the estimated means. On the right, convergence of the estimated standard-deviations. A likelihood evaluation refers to a likelihood calculation.

In Figure 3.6, we can notice that when the optimization terminates, the parameters have converged. It is also possible to remark that after about 400 evaluations, the estimation of the means has almost converged (exploration is over), while it takes longer (another 200 evaluations) for the standard-deviations to reach the MLE estimate $\widehat{\boldsymbol{\Psi}}$, showing the faster convergence of the mean parameters with respect to the standard-deviations. To conclude this analysis, the variability of the calibration can be examined. To carry out this examination, the minimization of the likelihood is performed for different initialization points that respectively correspond to $\nu \times \mathbf{1}_{\mathbb{R}^9}$ with $\nu \in \{0.15, 0.40, 0.65, 0.90\}$ (we limit the inquiry to 4 starts because of the computational costs). Then, the relative difference δ between the values of the population parameters for the best calibration $\widehat{\boldsymbol{\Psi}}_{\text{best}}$ (in terms of likelihood) and the three others $\widehat{\boldsymbol{\Psi}}$, defined as $\delta(\widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Psi}}_{\text{best}}) := 100 \times \frac{|\widehat{\boldsymbol{\Psi}} - \widehat{\boldsymbol{\Psi}}_{\text{best}}|}{\widehat{\boldsymbol{\Psi}}_{\text{best}}}$, are computed. They are then averaged over the calibrations that do not refer to the best one and reported for the 1st repetition of the calibration process in Table 3.4.

Table 3.4: Averaged relative difference δ in % of the mean parameters for different initializations for the first repetition of the calibration process. The random seed remains the same for all calibrations.

Population Parameter	μ_{E_0}	$\mu_{y_{0s}}$	$\mu_{y_{cs}}$	μ_{d_c}
Averaged relative difference [%]	0.0104	0.0413	0.0687	0.0435

Population parameter	sd(E_0)	sd(y_{0s})	sd(y_{cs})	sd(d_c)	ω
Averaged relative difference [%]	0.235	66.0	0.142	0.165	0.0395

Table 3.4 shows that each of the 4 calibrations for both repetitions estimate roughly the same set of population parameters and residual model, except the standard-deviation of the damage threshold. From these indicators, we can say that the chosen algorithm allows for a proper estimation of the population parameters $\boldsymbol{\Pi}$ and that its error is not mainly caused by the minimization process. The behavior of the standard-deviation of the damage threshold suggests that the objective function is not sensitive to this parameter, which is consistent with the low sensitivity of the model output with respect to y_{0s}, making imprecise the estimation of the corresponding mean and standard-deviation. This point is developed further in the next section, which presents the calibration of the population parameters.

3.4.1.3 Calibration of the population parameters

To study the results, we now investigate the calibration of the $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ through the estimation of the population parameters (means & standard-deviations) reported in Table 3.5.

Table 3.5: Averaged calibrated population parameters $\widehat{\boldsymbol{\Pi}}_{\text{mean}}$ and their associated averaged error $\mathcal{E}_{\text{mean}}(\widehat{\boldsymbol{\Pi}})$ in % over 20 repetitions of the calibration process with samples of 10 specimens each. The coefficients of variation COV in % are indicated between brackets. All model parameters are assumed independent.

	E_0 [GPa]	y_{0s} [kPa]	y_{cs} [MPa]	d_c [-]
Exact means	189	25.0	2.24	3.77
Averaged calibrated means	189 (3.60)	24.0 (14.6)	2.24 (5.64)	3.85 (6.57)
Averaged errors on calibrated means [%]	2.97 (68.6)	12.2 (64.5)	4.62 (70.2)	5.34 (86.0)
Exact standard-deviations	17.9	8.00	0.439	0.893
Averaged calibrated standard-deviations	15.9 (41.9)	3.96 (88.5)	0.502 (32.9)	0.844 (16.7)
Averaged errors on calibrated standard-deviations [%]	30.7 (77.7)	56.9 (61.6)	30.5 (87.1)	13.9 (65.8)

To start the analysis, let us first concentrate on the mean parameters, for which there is a good agreement between the exact and calibrated values, with levels of error smaller than 10%, except for the damage threshold. This is consistent with the errors committed on the individual parameters reported in Table 3.3, where the individual thresholds are estimated with larger error compared to the others. Its standard-deviation exhibits the same behavior as its average counterpart compared to the other parameters. Note that, generally, the mean parameters are better calibrated than the associated standard-deviations. For instance, the averaged relative error on μ_{d_c} increases from 5.34% to 13.9% for $\text{sd}(d_c)$. Higher increases can be remarked for the other parameters. This seems to be a general behavior that can be explained by structural reasons developed in the next paragraph.

The standard-deviations are estimated with larger errors with respect to the mean parameters. One possibility that can be raised is that the error on the individual parameters has a greater effect on the standard-deviations than on the mean parameters. This comes from the fact that the error on the estimation of the standard-deviations combines the errors on the individual parameters but also those on the mean parameters. Another reason is that standard-deviations parameters require more samples to converge to their estimate compared to the mean parameters (further discussed in Section 3.4.3). In fact, with few individuals (10 in the present case), it suggests that the likelihood is more sensitive to the mean parameters than to the standard-deviation parameters. This comes from the fact that the mean vector indicates the position of the most likely parameter, and nothing else but a decrease of the likelihood happens if it is poorly estimated, while the standard-deviations characterize a dispersion that is difficult to estimate with only 10 realizations. To illustrate this point, a rough sensitivity analysis can be conducted. It is carried out by computing the negative log-likelihood for population parameters that belong to $[0.9\widehat{\boldsymbol{\Psi}}, 1.1\widehat{\boldsymbol{\Psi}}]$, sampled with 50 points regularly spaced for the 1^{st} repetition. All but one parameter are held fixed to their MLE estimate. The negative log-likelihoods are scaled here by the optimal negative log-likelihood value.

Figure 3.7 shows that the standard-deviations do not seem to influence the likelihood as much as the means. Consequently, it is not surprising for them to exhibit higher estimation errors. This can be related with the convergence of the objective function value with respect to the convergence of the population parameters. Indeed, in Figure 3.5, the objective function reaches convergence after 400 opposite log-likelihood function evaluations, that is to say approximately

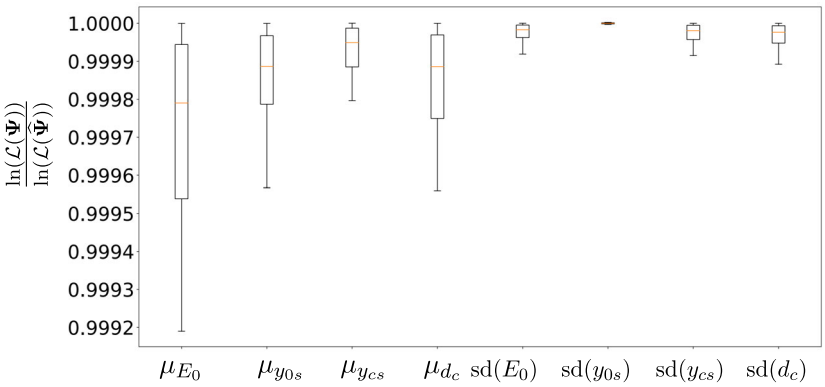


Figure 3.7: Relative variation of the negative log-likelihood with respect to the variation of each population parameters taken one by one. The negative log-likelihood value is computed for 50 points regularly spaced sampled within $[0.9\widehat{\boldsymbol{\Psi}}, 1.1\widehat{\boldsymbol{\Psi}}]$.

when the mean parameters have met a stationary behavior in Figure 3.6. Overall, we can see through the range of the y-axis that a change of 10% of the population parameters has a minor impact on the likelihood (at most to 0.01%), suggesting a high level of uncertainty on the population parameters. Table 3.5 and Figure 3.7 provide consistent results as the most sensitive parameters of Figure 3.7 correspond to the parameters calibrated with smallest level of errors. Notice, as Table 3.3 shows, that the standard-deviation of E_0 is estimated with a larger error than the standard-deviation of d_c while the individual moduli are estimated with errors comparable to those of the individual saturations. Still, E_0 is more concentrated around its mean than d_c. Indeed, the ratio of the exact standard-deviation divided by the exact mean reaches 9.50% for E_0 and 23.6% for d_c, lots of individual moduli will give the same model error. This effect has a small impact on the mean parameter because it is a sensitive position parameter. However, the standard-deviation, which is relatively small compared to the mean and indicates the dispersion of the model parameters, depends on the position of the individual parameters with respect to the mean value. The error on the standard-deviation of y_{cs} seems consistent with level of errors committed on the estimation on the individual kinetics. The same occurs for the standard-deviation of y_{0s} which do not change the likelihood function and is consistent with the associated level of errors. The calibration of the marginals is illustrated for one of the repetitions (the 1$^{\text{st}}$ repetition) in Figure 3.8 and depicts especially the concentration of the individual thresholds y_{0s} around their associated mean and the incomplete coverage of the 95% interval with only 10 individuals.
In Figure 3.8, the effect of the regularization effect of the term $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})$ in the likelihood (see Eq. (3.5) which reduces to $\mathbf{b}_i^\top \boldsymbol{\Sigma}^{-1}\mathbf{b}_i$ for Gaussian distributed parameters) is illustrated, in particular for the damage threshold (top right in Figure 3.8). The impact of this regularization is reflected in the parameter space by the concentration of the individual threshold around their mean value, a phenomenon that is sometimes called shrinkage and quantified in the next section. This occurs specifically for this parameter because of the low sensitivity of the model output with respect to the damage threshold, as can be seen on the top right of Figure 3.8. It also shows that the marginals of E_0 and d_c are better estimated than the one of y_{cs}, accordingly to Table 3.5. The same figure also displays the calibrated individual parameters. The values of the individual parameters are all different, which expresses the variability of the material specimen. This is a key feature with mixed-effects, contrary to the individual approaches studied in Chapter 1 where all individuals are supposed identical. It also associates the calibrated individual parameters to their exact counterparts, hinting for an acceptable estimation of the individual parameters (developed in the next Section). Still, on average, its mean parameter is close to the target value, contrary to its standard-deviation that suffers from this shrinkage, discussed further below. From these considerations, it follows that the calibrated population parameters can be considered as properly estimated, provided the inherent limitations of the number of individuals $n = 10$, the difficulties brought by the low sensitivity of the damage threshold and of the damage evolution

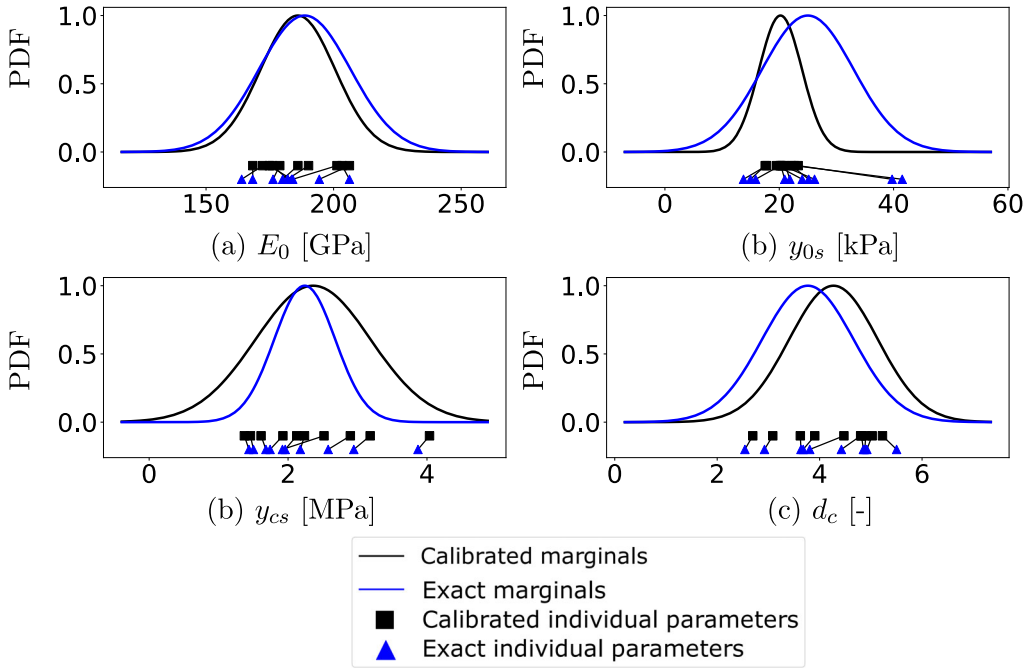


Figure 3.8: Exact and calibrated marginals along the E_0 axis (a), the y_{0s} axis (b), the y_{cs} axis (c) and the d_c (d) axis for the 1$^{\text{st}}$ repetition. Exact and calibrated individual parameters are also plotted.

celerity to a certain extent, and the concentration of the Young's moduli around its average. The impact on the estimation of the population of the number of individuals is investigated later.

3.4.1.4 Calibration of the individual parameters

Mixed-effects calibration is based on the calibration of the individual parameters. Thus, to assess the calibration of the population parameters, it is necessary to examine the estimation of the individual parameters. To analyze this estimation, it is possible to investigate the level of errors for the different model parameters. Another indicator that can be used is the η-shrinkage that aims to measure whether the data is informative enough or not to estimate precisely the individuals parameters. The η-shrinkage is computed using the formula that involves the standard-deviations, provided that the optimization variables refers to the standard-deviations rather than the variances and is expressed as $\text{sh}(\eta) := 1 - \frac{\sqrt{\widehat{\mathbb{V}}(\widehat{\theta}_{il})}}{\sqrt{\mathbb{V}(\theta_{il})}}$, with $\sqrt{\widehat{\mathbb{V}}(\widehat{\theta}_{il})}$ the empirical standard-deviation of the calibrated individual parameters and $\sqrt{\mathbb{V}(\theta_{il})}$ the standard-deviation calibrated by maximum likelihood [Lavielle, 2014]. The smaller the $\eta-$shrinkage, the closer $\sqrt{\widehat{\mathbb{V}}(\widehat{\theta}_{il})}$ is to $\sqrt{\mathbb{V}(\theta_{il})}$, indicating that individual experiments properly cover the range of variation of the parameters and the regularization term has less effect. Both indicators are reported respectively in Tables 3.6 and 3.7.

Table 3.6: Averaged relative errors $\mathbf{e}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_{10})$ in % on the 10 individual parameters averaged over 20 repetitions of the calibration process. The corresponding coefficients of variation COV in % are indicated in brackets. All model parameters are assumed independent.

	E_0	y_{0s}	y_{cs}	d_c
Averaged relative error [%]	2.89 (27.6)	27.1 (51.1)	8.55 (19.2)	4.31 (23.1)

Table 3.7: Averaged $\eta-$shrinkages in % for 10 individuals over 20 repetitions of the calibration process. The coefficients of variation COV in % are indicated between brackets. All model parameters are assumed independent.

	E_0	y_{0s}	y_{cs}	d_c
$\eta-$shrinkage [%]	14.6 (1.1)	60.8 (52.4)	9.33 (62.9)	2.77 (53.0)

Table 3.6 confirms that, over the 10 individuals and the 20 repetitions, all parameters can be considered to be are accurately calibrated except the damage threshold, given the presence of noise and of the regularization term. In fact, as mentioned earlier, the data is not informative enough to determine precisely y_{0s}, which is confirmed by the $\eta-$shrinkages computed from the estimated individual parameters and detailed in Table 3.7. Furthermore, though it is much smaller than for the damage threshold, the η-shrinkage for E_0 is greater than those for either y_{cs} and d_c. One possible explanation is the small variation of this parameter with respect to its mean value, which makes it more difficult to discriminate between the model parameters and thus decreasing the strength of the fitting term with respect to the regularization term within function $g_i(\cdot)$ in Eq. (3.5). The agreement between the exact and calibrated individual parameters is also illustrated in Figure 3.9 that depicts the exact individual parameters versus the calibrated ones. If they are suitably identified, they should be close to the first bisector (the gray line in Figure 3.9), as it occurs for y_{cs}, d_c and to a lesser extent, E_0. On the contrary, the estimations of the individual thresholds are poor, as discussed earlier (see Table 3.6).

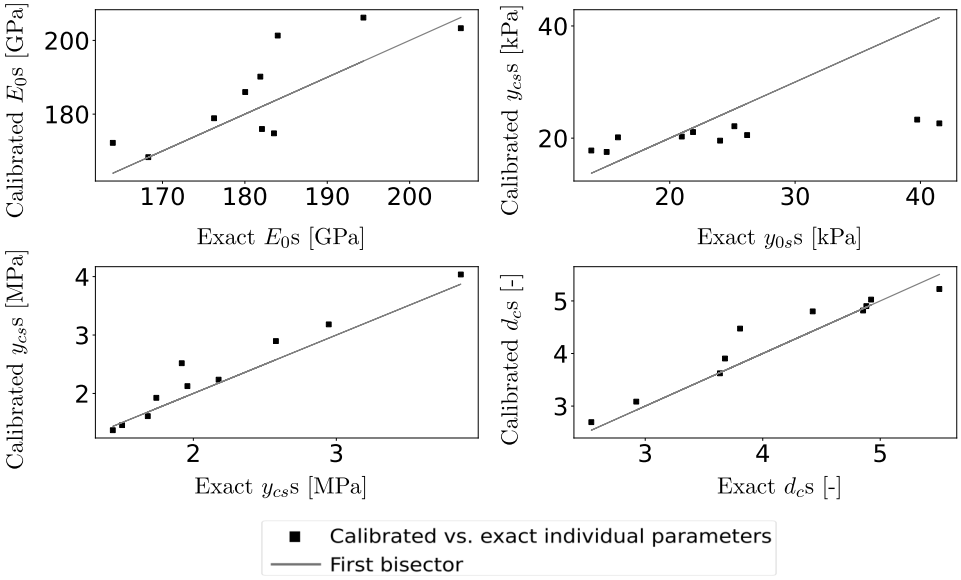


Figure 3.9: For the 1st repetition, exact versus calibrated individual parameters for the E_0 (top left), y_{0s} (top right), y_{cs} (bottom left) and d_c (bottom right).

3.4.1.5 Calibration of the specimen curves

In addition to the parameter space, the calibration of the individuals should also be checked in model space. A simple test consists into comparing the calibrated models to both the exact and noisy data, as illustrated for the 10 specimens involved in the calibration process in Figure 3.10.

Here, we can notice that, for all individuals, the models predictions computed with the calibrated individual parameters accurately fit the target strain-stress curves: the averaged distance between the exact and calibrated curves is below 2 MPa (1.76 MPa on average for a coefficient of variation of 14%), a small value in comparison to the target stress which exceeds 100 MPa. The same conclusions arise from Figure 3.11 and from Figure 3.10 which displays for the 1st repetition the difference between the exact and calibrated curves alongside the same between

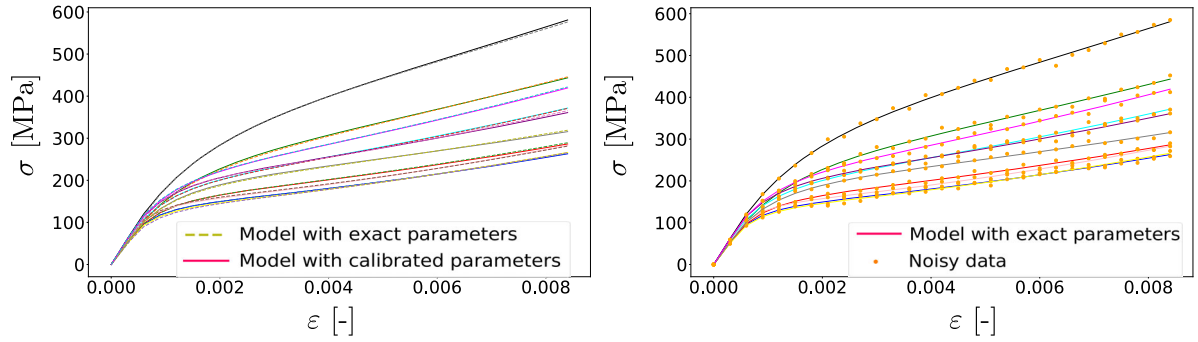


Figure 3.10: Calibration of the individuals involved in the calibration process. On the left, stress-strain curves for the exact and calibrated individual parameters. On the right, stress-strain curves for the calibrated individual parameters and noisy data.

the noisy and calibrated curves. The difference between either the exact and calibrated curves or between the noisy and calibrated curves remain below 10 MPa for most observation points, that is to say below 2% of the maximum stress output that is about 500 MPa.

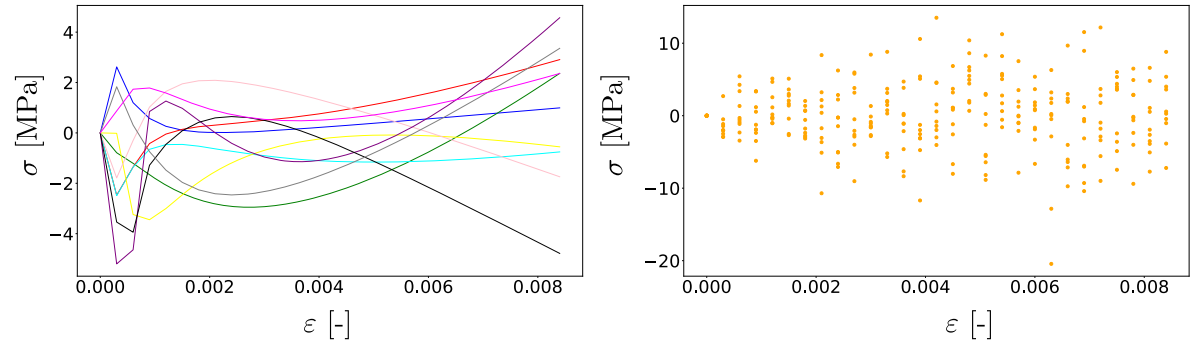


Figure 3.11: Calibration of the individuals involved in the calibration process for the 1st repetition. On the left, difference between the exact and calibrated stress-strain curves. On the right, difference between the calibrated stress-strain curves and the noisy data.

3.4.1.6 Evaluation of the residual model

The calibration has been carried out by assuming that the residuals, *i.e.*, the difference between the noisy data and the model output, is Gaussian distributed with zero mean and unknown variance labeled ω_i^2 which needs to be calibrated. Furthermore, it is supposed to be homoscedastic over all the individuals which gives a unique parameter for the residual model. To complete the investigation process, this model should be validated. It is possible first to focus on the empirical residuals of the calibration curves. In Figure 3.12, they are depicted for two experiments (the 2nd and the 6th) of the 1st repetition, or rather the standardized residuals $\widehat{\mathbf{r}}_{S,i}$ defined as $\widehat{\mathbf{r}}_{S,i} := \frac{\mathbf{y}_i - \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)}{\widehat{\omega}}$ (that should thus follow $\mathcal{N}(\mathbf{0}_{\mathbb{R}^{N_i}}, \mathbf{I}_{N_i})$).

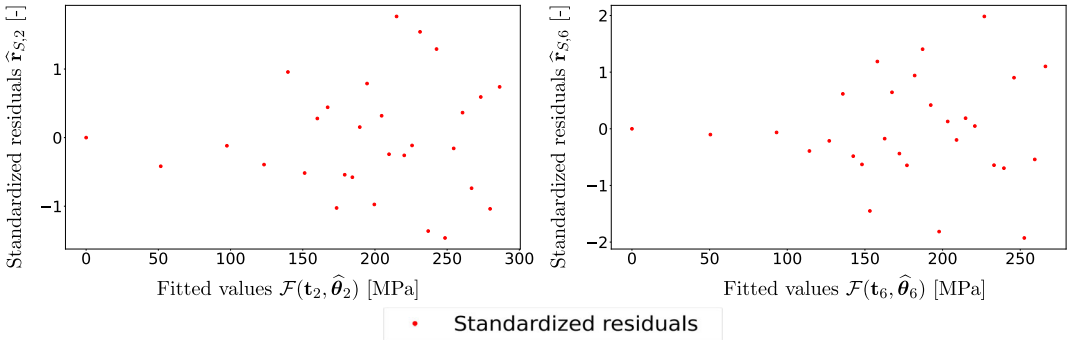


Figure 3.12: For the 1st repetition, standardized residuals for two individuals randomly selected (the 2nd and the 6th).

The residuals displayed in Figure 3.12 do not seem to exhibit specific structure, suggesting that model bias is not significant. Note that the residual that corresponds to the first point in both graphs is always 0 due to the model formulation, whichever proposal of individual parameters is made. Note also that the residuals are centered, bounded between -2 and 2, which corresponds to the 95.4% confidence interval for a Gaussian distribution. The residuals seem to grow with the stress values which suggest their heteroscedastic nature (that is indeed the case). To improve the residual model, one possible assumption is to suppose heteroscedasticity between the linear and nonlinear regime. Given that the linear regimes is in fact composed of only one point provided the threshold values, a proper estimation of the variance of the noise for the linear part would be challenged as this variance gets more accurate as the number of points involved increases. We can also study the assumption of normality for the standardized residuals, which can be achieved using Quantile-Quantile plot as depicted in 3.13.

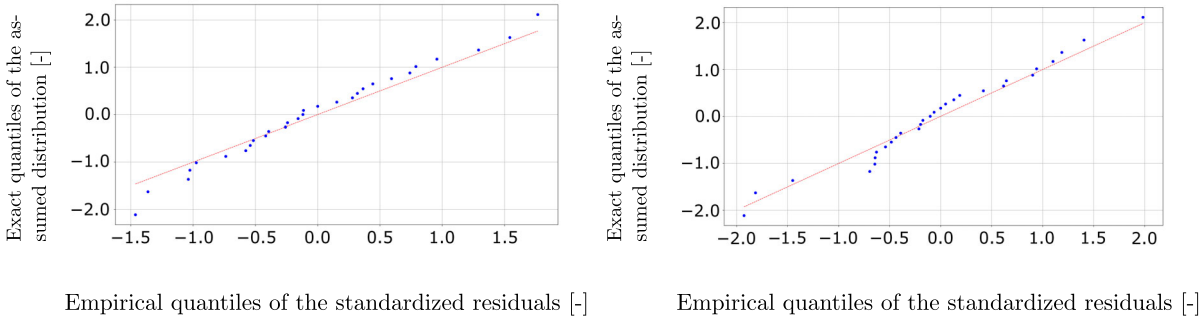


Figure 3.13: For the 1st repetition, normal QQplot of the standardized residuals for two experiments randomly selected: the 2nd on the left and the 6th on the right. In red, the first bisector on which data points (in blue) should be located.

The QQ-plots provided in Figure 3.13 do not seem to confirm the assumption of Gaussian distributed residuals, as points deviate from the first bisector. However, though it is not exact, this hypothesis does not appear to prevent a proper estimation of both the population and individual parameters, given the limitations coming from the formulation of the behavior law, the chosen parameters values and the number of available individuals. It may only slightly harden the calibration process but does not seem to introduce bias in the parameters estimate. This is consistent with the conclusions from [Demidenko, 2013] who states that an error in the residual model does not usually prevent a proper calibration, contrary to an erroneous choice of the model parameters distribution. Consequently, the assumption of homoscedastic Gaussian distributed residual will be kept for this part for the sake of simplicity. Note also that because of the formulation of the minimization problem to estimate the individual parameters, the residual may exhibit model bias. Indeed, the presence of the regularization term in Eq. (3.5) can prevent the model prediction for the $\widehat{\boldsymbol{\theta}}_i$s to match perfectly with the exact data.

These first results provide encouraging conclusions about the application of mixed-effects calibration for mechanical models. Yet, the minimization with the CMA-ES algorithm of the opposite log-likelihoods with the objective function evaluations at each iteration carried out simultaneously over 11 cores (the size of the population) Broadwell Intel® Xeon® CPU E5-260 v4 @ 2.20GHz takes about one hour. In fact, the computationally intensive phase of the calculation is the identification of the individual parameters, carried out so far with the CMA-ES algorithm. For instance, on average, a likelihood minimization requires about 10^8 model evaluations. Thus, to reduce significantly the computational time, the most efficient option is to reduce the numerical costs of this phase. Indeed, with more complex models and individuals, we may wonder whether the method would be able to get MLE estimates at reasonable computational costs. However, remember that the algorithm should still provide accurate estimates of the individual parameters. The choice of the minimization algorithm is studied in the following section.

3.4.2 Studying the choice of the auxiliary optimization algorithm

Recall that the objective of mixed-effects calibration is to adjust the population parameters to make the model parameters distribution best fit the empirical distribution of the individual parameters. Therefore, the chosen method should be able to identify properly the individual parameters on the different experiment repetitions. For the Laplace approximation, this is performed through the minimization of function $g_i(\cdot)$ defined in Eq(3.5). So far, we have chosen to achieve the auxiliary minimization with an evolutionary algorithm that enables both exploration and exploitation to ensure a proper identification of the individual parameters. Nevertheless, such a choice implies large computational costs, because it is necessary to test many candidates to find the optimal solution [He et al., 2023], which may prevent calibration of complicated models. In this section, we aim to check whether the evolutionary algorithm can be replaced by a multi-start gradient descent. This choice combines the convergence speed of the gradient descent alongside an exploration with different initializations to lower the risk of being trapped into local minimums. The calibrations with either the CMA-ES algorithm or a multi-start gradient algorithm are compared in terms of accuracy for the estimation of the full population parameters and computational costs. The assumptions on the model parameters to calibrate remain the same. The gradient algorithm is the Sequential Least Squares Programming (SLSQP) [Johnson, 2007] algorithm based on the library NLopt included in the OpenTURNS library [Baudin et al., 2016]. To keep computational costs reasonable, 10 different starts are considered. 9 of them are determined from a Latin Hypercube Sampling, the last corresponding to the population average. Similar settings for the monitoring of convergence from the CMA-ES algorithm are chosen. The gradient of function $g_i(\cdot)$ is estimated by finite differences with default step for its calculation. The maximum number of model evaluations is set to 1,000. The same synthetic data is involved in both cases, such as the global optimization of the likelihood that remains achieved with CMA-ES algorithm (with the same optimization settings). Successively are studied the calibration of the individuals with the SLSQP algorithm, the consequences on the estimation of the population parameters and the comparison of the computational costs. The different steps are described in Figure 3.14.

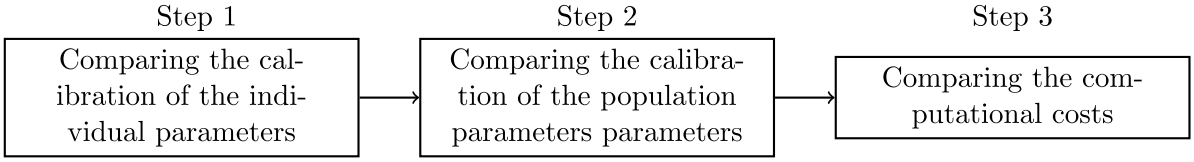


Figure 3.14: Flowchart of the different steps for the analysis of the impact of the choice of the algorithm for the auxiliary minmization.

3.4.2.1 Calibration of the individuals with the SLSQP algorithm

The only difference between this section and the previous one resides in the determination of the individual parameters. Thus, before performing mixed-effects calibration for the ODM with the same data, we aim to make sure that minimization of function $g_i(\cdot)$ occurs without issues. This can be studied by depicting the objective function values throughout the iterations as proposed in Figure 3.15. In this figure, at a given iteration, only the best objective value out the 10 is plotted.

In Figure 3.15, for both proposal of full population parameters, the minimization of function $g_i(\cdot)$ seems to be carried out without problems. This is of prime importance as otherwise, numerical instabilities on the minimization of the global negative log-likelihood could occur. To validate this method, it is import to study to what extent it can approximate the individual parameters with respect to the choice of the full population parameters. To achieve this, we uniformly draw samples (1,000) within the bounds defined in Table 3.2. For each of these samples, the individual parameters are estimated and the relative error is computed from Eq. (3.6). Furthermore, to examine the impact of the choice of the auxiliary algorithm, we estimate the averaged relative

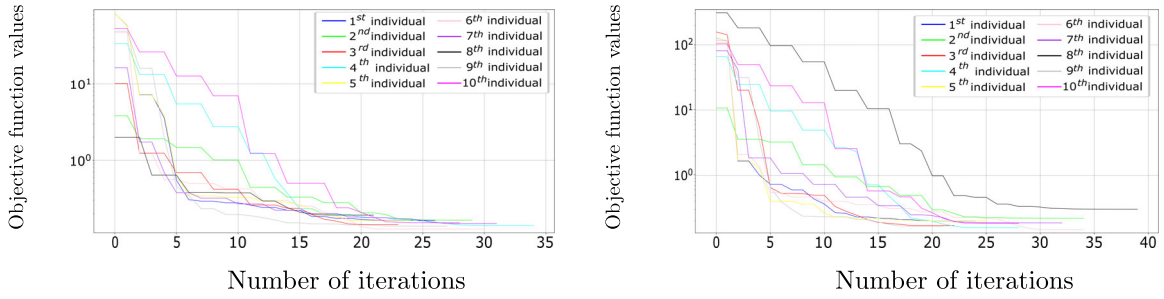


Figure 3.15: For the calibration with the gradient descent, objective function values $g_i(\cdot), i \in [\![1, 10]\!]$ for the 1st repetition. On the left, $\boldsymbol{\Psi} = \boldsymbol{\Psi}_0$ (the initialization point) and on the right, $\boldsymbol{\Psi} = \widehat{\boldsymbol{\Psi}}$ (the MLE estimate). At a given iteration, only the best objective value out of the 10 restarts is plotted.

difference between the individual parameters calibrated with both algorithms $\delta(\boldsymbol{\theta}_{i,\text{CMA}}, \boldsymbol{\theta}_{i,\text{SLSQP}})$ that is expressed as the averaged relative error defined in Eq. (3.6), except that $\theta_{il,\text{exact}}$ is replaced by its estimation with the CMA-ES algorithm and θ_{il} by its estimation with the SLSQP algorithm. This test is carried out with the data of the 1st repetition for the same proposal of $\boldsymbol{\Psi}$ used for Table 3.3.

Table 3.8: For the 1st repetition, ability of the SLSQP algorithm to determine the individual parameters for 1,000 proposals of full population parameters $\boldsymbol{\Psi}$ sampled between the bounds defined in Table 3.2. Two indicators are studied: the averaged relative difference between the individual parameters calibrated with both algorithms $\delta(\boldsymbol{\theta}_{i,\text{CMA}}, \boldsymbol{\theta}_{i,\text{SLSQP}})$ and the average relative errors $\mathbf{e}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_{10})$, both in %.

	E_0	y_{0s}	y_{cs}	d_c
Relative difference $\delta(\boldsymbol{\theta}_{i,\text{CMA}}, \boldsymbol{\theta}_{i,\text{SLSQP}})$ [%]	3.42×10^{-3}	2.78×10^{-2}	8.02×10^{-3}	4.59×10^{-3}
Averaged relative error on the individual parameters using the SLSQP algorithm [%]	10.6	55.7	20.7	13.8

Table 3.8 shows that the SLSQP algorithm provides the same estimation of individual parameters as the CMA-ES algorithm (relative difference lower than 10^{-3}%), and thus reaches the same level of accuracy for different shapes of material variability (see Table 3.3), which allowed in the earlier section for a proper estimation of the population parameters. Furthermore, the default settings for this algorithm can be retained as with them, there is no deterioration of the estimation of the individual parameters compared to the CMA-ES algorithm. The robustness of the SLSQP algorithm to carry out this minimization validates its use to perform the auxiliary optimization. Before proceeding to the calibration of the population parameters, the need of a multi-start algorithm should be underlined. To demonstrate this, we present the results of this auxiliary minimization for the initialization points of the LHS sample and the mean vector given $\boldsymbol{\Psi} = \widehat{\boldsymbol{\Psi}}$. The result of the corresponding minimization is underlined in green in Figure 3.16.

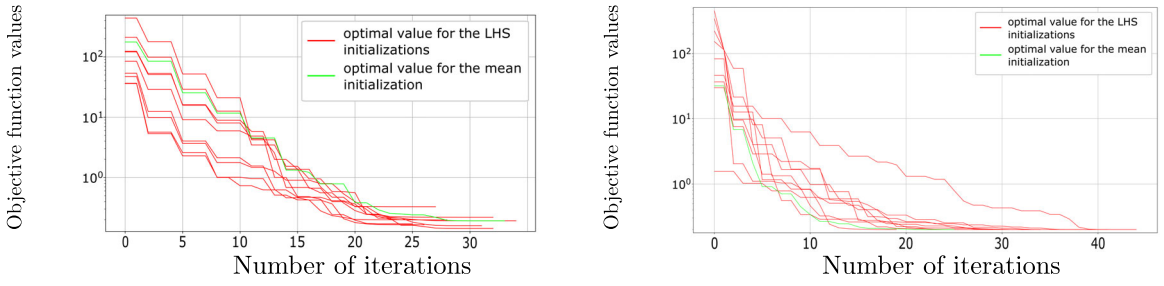


Figure 3.16: Objective function values $g_i(\cdot), i \in \{2, 6\}$ for the second (on the left) and sixth individual (on the right) randomly selected of the 1st repetition given $\boldsymbol{\Psi} = \widehat{\boldsymbol{\Psi}}$ (the MLE estimate) for the different points of initialization of the LHS sample (red curves) and the mean vector (green curve). Each curve refers to a distinct optimization attached to one of the selected starting points.

In Figure 3.16, the minimization with the estimated mean vector as the initialization point does not turn out to be always the most relevant to estimate the individual parameter as it does not achieve the best minimum. Given that it is not possible to tune the auxiliary optimization during the likelihood optimization, we decide to keep this multi-start for the identification of the individual parameters that strengthens the identification of the individual parameters. In addition, it implies that the calculation of the gradient of function $g_i(\cdot)$ is not particularly flawed as the minimization converges properly and identifies properly the individual parameters. To terminate the verification, it remains to investigate whether the use of the SLSQP introduces numerical issues in the minimization of the likelihood. This point can be studied by depicting the opposite log-likelihood values throughout the optimization process in Figure 3.17.

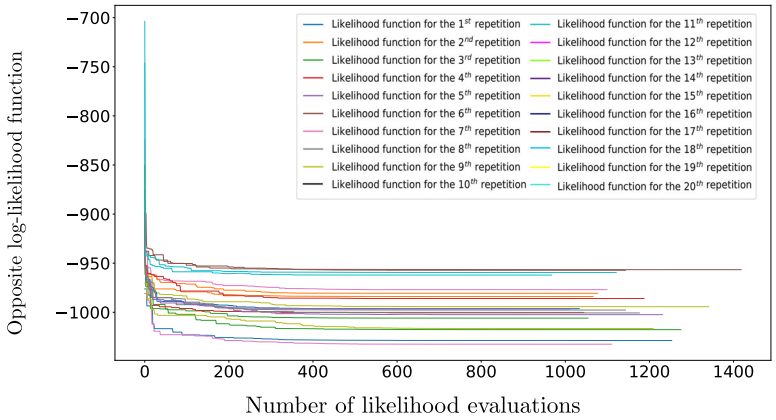


Figure 3.17: For all repetitions, best likelihood value at each evaluation.

Figure 3.17 displays the levels of likelihood taken throughout the optimization process for the 20 repetitions. In fact, the important point here to notice from both Figures 3.5 and 3.17 is that changing the auxiliary optimization algorithm does not alter the convergence properties of the likelihood minimization. The calibration of the population parameters can be now studied.

3.4.2.2 Calibration of the population parameters.

We can now check the results of mixed-effects calibration for the ODM with the multi-start SLSQP for the auxiliary optimization. The first step consists in checking the estimation of the population parameters by comparing among others the errors committed on the population parameters with both the SLSQP and CMA-ES algorithm which are reported in Table 3.9. The computational costs for both methods are studied next.

Table 3.9: Comparison of the averaged error on the population parameters $\mathcal{E}_{\text{mean}}(\widehat{\mathbf{\Pi}})$ in % over 20 repetitions of the calibration process with samples of 10 specimens each using both CMA-ES and SLSQP algorithms for the auxiliary optimization. The coefficients of variation COV in % are indicated between brackets.

	E_0	y_{0s}	y_{cs}	d_c
Averaged errors on calibrated means with CMA-ES algorithm [%]	2.97 (68.6)	12.2 (64.5)	4.62 (70.2)	5.34 (86.0)
Averaged errors on calibrated means with SLSQP algorithm [%]	2.92 (66.0)	11.7 (68.8)	4.58 (72.3)	5.51 (81.3)
Averaged errors on calibrated standard-deviations with the CMA-ES algorithm [%]	30.7 (77.7)	56.9 (61.6)	30.5 (87.1)	13.9 (65.8)
Averaged errors on calibrated standard-deviations with the SLSQP algorithm [%]	31.6 (80.2)	53.0 (69.2)	32.0 (82.3)	13.8 (67.5)

In Table 3.9, replacing the CMA-ES by a multi-start gradient descent does not seem to downgrade much the estimates of the population and individual parameters. In fact, the relative

errors committed considering the two auxiliary optimization algorithms are similar, whichever calibrated parameters is considered. This is illustrated on the left of Figure 3.18, where the marginals of d_c calibrated using both the SLSQP (in green) and the CMA-ES algorithm (in black) for the auxiliary optimization are compared to their exact counterpart. It also performs the same comparison for the individual parameters. Since the marginal distributions are jointly calibrated, a relevant criterion to measure the accuracy of the calibration that accounts for the complete joint parameter distribution is the Kullback-Leibler (KL) divergence on which we rely. For the reasons exposed earlier, y_{0s} cannot be properly learned here and is omitted in the KL calculation. The same occurs for the individual parameters as shown in Table 3.10.

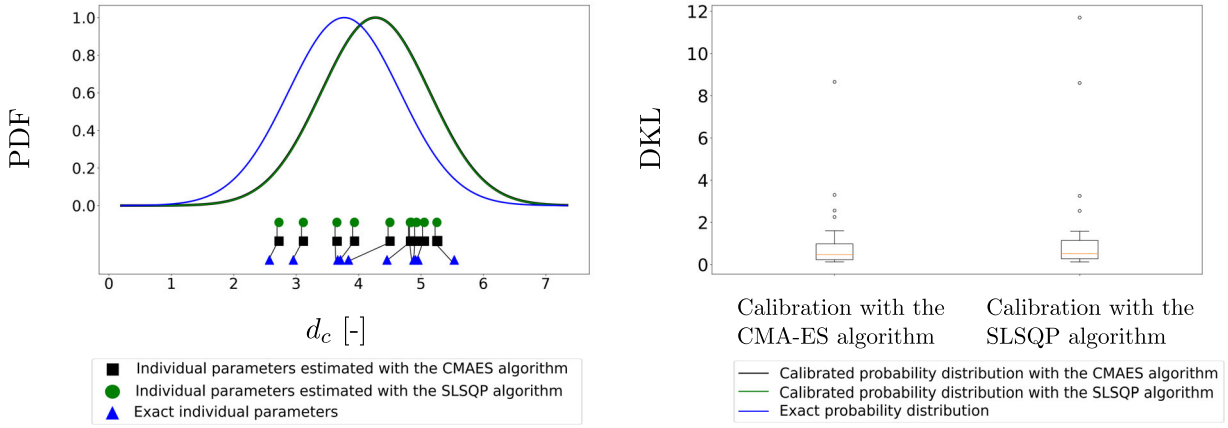


Figure 3.18: Illustration of the similarity of the results with respect to the choice the algorithm for the auxiliary optimization. On the left, for the first repetition, exact and calibrated marginals (individual parameters) of the damage saturations for SLSQP and the CMA-ES algorithm. On the right, KL divergence between the exact and estimated distribution for calibrations carried out with SLSQP and the CMA-ES algorithms for the auxiliary optimization.

Table 3.10: Comparison of the relative errors $\mathbf{e}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_{10})$ in % on the 10 individual parameters averaged over 20 repetitions of the calibration process using both CMA-ES and SLSQP algorithms for the auxiliary optimization. The corresponding coefficients of variation in % are indicated in brackets. All model parameters are assumed independent.

	E_0	y_{0s}	y_{cs}	d_c
Averaged relative error on the individual parameters using the CMA-ES algorithm [%]	2.89 (27.6)	27.1 (51.1)	8.55 (19.2)	4.31 (23.1)
Averaged relative error on the individual parameters using the SLSQP algorithm [%]	2.79 (27.7)	23.6 (40.0)	8.36 (21.8)	4.19 (24.2)

Both algorithm gives similar results in terms of calibration of the model parameters distribution. This behavior is not surprising provided that both algorithms give almost the same estimation of the individual parameters (see Table 3.8) and that the population parameters are calibrated from these estimations. From the previous indicators (among which a similar estimation of the population and individual parameters), it arises that performing the auxiliary minimization with either the CMA-ES or the SLSQP algorithm provide the same conclusion: the minimization of the opposite log-likelihood minimization converges to a minimum that seems to be global. Then, it appears that we can validate the use of the SLSQP algorithm for the auxiliary optimization. We can now compare the computational cost of each solution.

3.4.2.3 Studying the impact of the auxiliary optimization on the computational costs

The numerically costly phase of the likelihood computation is the resolution of the minimization problem described in Eq. (3.6). Remember that the default settings kept in this investigation

provided accurate estimates of the population parameters despite of the numerical challenges discussed above. For both algorithms, at most 1,000 model evaluations were allowed. To study the impact of the algorithm on the computational costs, two indicators can be examined: the total number of model evaluations (through the maximization of the likelihood) and the number of iterations of the algorithm for the minimization of the likelihood considering both algorithms for the auxiliary optimization. Indeed, the first is of prime interest because it refers to the expensive part of the likelihood computation while the second aims to check to what extent the convergence of the minimization is altered. In addition, in both cases, the minimization of the global opposite log-likelihood defined in Eq. (3.8) is carried out with the CMA-ES algorithm and the same settings as earlier are chosen for both minimizations. The settings of the auxiliary optimization have been provided in Section 3.3.2.2 when the CMA-ES algorithm is used and at the beginning of this section when a multi-start gradient is used. Finally, note that both algorithms provide comparable results which can be considered as satisfying given the modeling assumptions, the number of available individuals and the values of the population parameters.

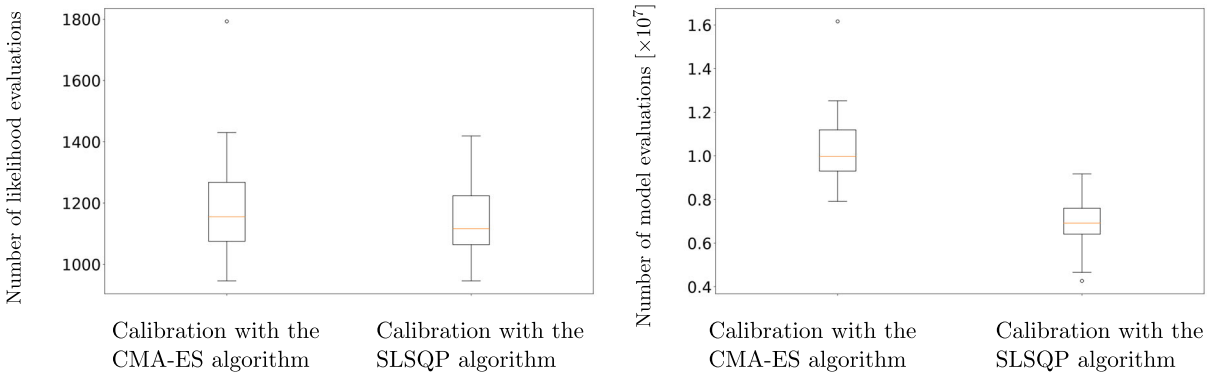


Figure 3.19: Comparison of the computational costs with respect to the choice the algorithm for the auxiliary optimization. On the left, likelihood evaluations and on the right, number of ODM model evaluations with both the SLSQP and the CMA-ES algorithms.

Figure 3.19 displays the number of model and likelihoods evaluation throughout the optimization process using both the SLSQP and CMA-ES algorithms. It shows that, even if the minimization with SLSQP algorithm tends to call the likelihood function more than when the auxiliary optimization is performed with the CMA-ES algorithm, there is a net gain of about 25% in terms of model evaluations which represents a significant improvement for the speed of the calibration. However, even with the SLSQP algorithm, the number of model evaluations remains large (about 5×10^6), and calibrating complicated model may require to set up alternative strategies such as metamodeling [Forrester et al., 2008].

In this section, we proposed to perform the auxiliary minimization with a gradient based algorithm which enables to speed up the minimization of the likelihood without significant loss of accuracy. It is possible now to study more in detail different settings such as the number of individuals.

3.4.3 Investigating the influence of the number of individuals

So far, 10 individuals have been involved in the calibration process. Though this number of specimens can be considered as large in the field of mechanics, it is considered as a very restricted number of samples in statistics, especially for the estimation of a multivariate distribution of dimension 4. This lack of specimens partly explains the error on the population parameters. To examine this issue, the ODM model is calibrated with mixed-effects considering 20 and 50 individuals. In fact, the 10 individuals of the previous sections are a subset of the 20 as the 20 are a subset of the 50. The different assumptions and settings remain identical as those used earlier. Thus, we can compare the calibration results to evaluate the consistency of the method,

especially to answer the following question: to what extent increasing the number of specimens decreases the error/variability of the estimates? The impact of the number of individuals is studied first on the estimation error of the population parameters, then on their uncertainty, and finally on the computational costs.

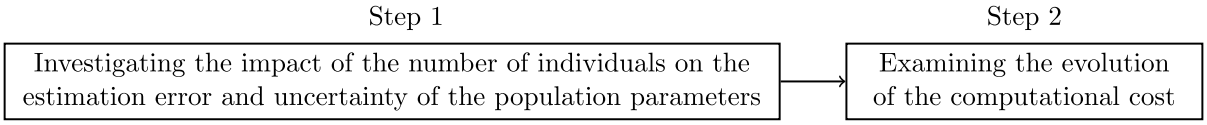


Figure 3.20: Flowchart of the different steps for the evaluation of the consequences of the number of individuals on the mixed-effects calibration results.

3.4.3.1 Characterizing the convergence of the population parameters

To start this study, the estimation errors on the population parameters for 10, 20 and 50 individuals are reported in Table 3.11.

Table 3.11: Averaged error $\mathcal{E}_{\text{mean}}(\widehat{\mathbf{\Pi}})$ in % for the population parameters over 20 repetitions of the calibration process with samples of 10, 20 and 50 specimens each. The coefficients of variation COV in % are indicated between brackets. All model parameters are assumed independent.

	E_0	y_{0s}	y_{cs}	d_c)
Averaged error on mean parameters for 10 individuals [%]	2.92 (66.0)	11.7 (68.8)	4.58 (72.3)	5.51 (81.3)
Averaged error on mean parameters for 20 individuals [%]	1.90 (89.8)	7.16 (86.1)	3.80 (89.8)	3.14 (56.2)
Averaged error on mean parameters for 50 individuals [%]	1.42 (66.8)	5.15 (71.0)	2.03 (59.6)	2.42 (73.3)
Averaged error on standard-deviations parameters for 10 individuals [%]	31.6 (80.2)	53.0 (69.2)	32.0 (82.3)	13.8 (67.5)
Averaged error on standard-deviations parameters for 20 individuals [%]	18.6 (79.6)	50.8 (72.0)	21.7 (106)	8.02 (89.6)
Averaged error on standard-deviations parameters for 50 individuals [%]	9.10 (83.9)	31.0 (84.6)	14.1 (71.5)	7.41 (60.6)

Tables 3.11 shows that adding individuals increases the accuracy of the results as it could be expected. Yet, the error on the population parameters does not decrease at the same speed for means and standard-deviation parameters. Indeed, with 10 individuals, the mean parameters are almost already converged (except for the damage threshold) while significant error can be committed for the estimation of the standard-deviations. With 20 individuals, the accuracy of the estimated standard-deviations increases by a significant amount (the error on the estimation of the standard-deviations decreases by 40% for all parameters except y_{0s}) and this effect is even more pronounced with 50 individuals where the maximum error reaches 15% for all parameters except y_{0s}. In addition, even if it is much slower than for the other parameters, there seems to be some convergence for the marginal of y_{0s}. Yet, contrary to the level of errors, the COVs do not seem to decrease (so to converge). One possible explanation is that, for some repetitions, some individuals can be outside the 99% credible interval (they refer to outliers) which is not surprising as in total, 1,000 individuals parameters have been generated. In fact, it shows that, even with a 50 individuals, it remains some sensitivity to the outliers which can flaw the estimation of the population parameters. To summarize Table 3.11, it seems possible to have a good approximation of the mean parameters when the 20 individuals appear to be enough to reach convergence. On the contrary, 20 individuals are required to get a proper estimation of the standard-deviations while with 50, a convergence behavior seems to be attained for all population parameters. It can be illustrated for one repetition over the marginal d_c as in Figure

3.21, that shows in particular, how the settlement of the support of the distribution allows a better calibration as the number of individual increases.

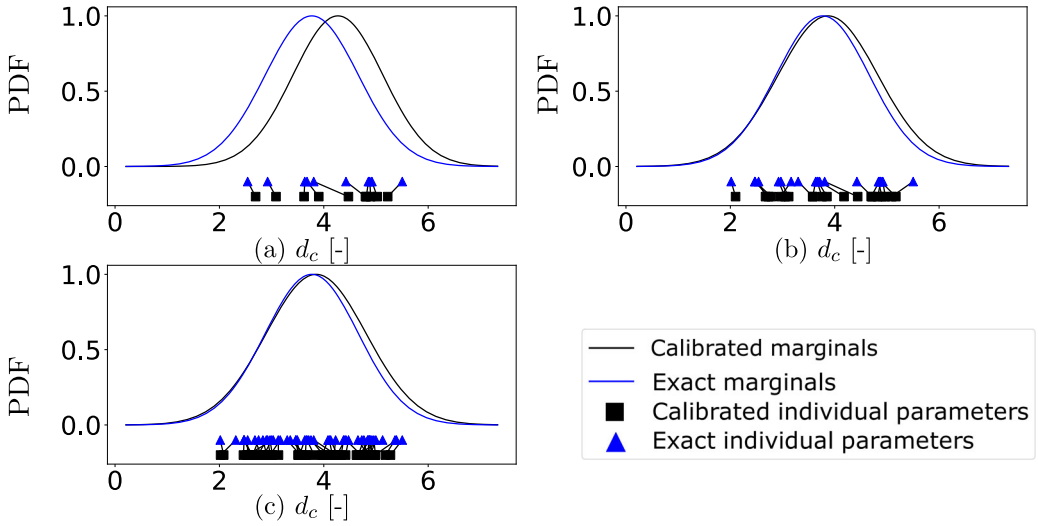


Figure 3.21: Illustration of convergence for mixed-effects calibration of the ODM model in parameters space with the exact and calibrated marginals along the d_c axis estimated with 10 (a), 20 (b) 50 individuals (c) for the 1st repetition. Individual parameters are also plotted.

We can also check, either in Figure 3.21 or Table 3.12, the estimation of the individual parameters is not downgraded for the different number of individuals considered here.

Table 3.12: Comparison of the relative errors $\mathbf{e}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_n)$ in % on the individual parameters averaged over 20 repetitions of the calibration process with 10, 20 and 50 individuals. The corresponding coefficients of variation in % are indicated in brackets. All model parameters are assumed independent.

	E_0	y_{0s}	y_{cs}	d_c
Averaged relative error on the individual parameters with 10 individuals [%]	2.79 (27.7)	23.6 (40.0)	8.37 (21.8)	4.19 (24.2)
Averaged relative error on the individual parameters with 20 individuals [%]	3.01 (18.2)	25.4 (17.9)	8.58 (14.9)	4.58 (17.8)
Averaged relative error on the individual parameters with 50 individuals [%]	2.91 (10.3)	24.7 (19.3)	8.21 (10.5)	4.38 (10.9)

This convergence can also be characterized using the KL divergence that takes into account the different dimensions together. Figure 3.22 summarizes the 20 repeated calibrations as boxplots of the KL divergences for 10, 20 and 50 individual specimens. It shows, when the number of specimens increases, the distribution is better calibrated. It also illustrates that up to a number of individuals, an asymptotic regime is met, and that, consequently, the improvement of the KL divergence (and consequently the gain in quality of the calibrated $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$) decreases. In fact, the results from both Table 3.11 and Figure 3.21 illustrate the same result: the convergence of the calibrated distribution with respect the number of individuals to the exact distribution.
These observations suggest that an asymptotic regime is met as the number of experiments increases and thus, precision gains are lower. Yet, the error criterion is not sufficient to characterize the convergence. Indeed, because of the finite number of individuals, the estimate of $\boldsymbol{\Psi}$ is flawed with uncertainties. To estimate these uncertainties, the asymptotic theory or bootstrap can be used as described in Section 2.4.1 of Chapter 2. A common limitation from both methods is that they are based on an asymptotic behavior, in other words, when the numbers of available specimens goes to infinity. With 10 individuals and to a lesser extent 20, this assumption is not met, and the results thus demand to be analyzed carefully. The asymptotic theory demands

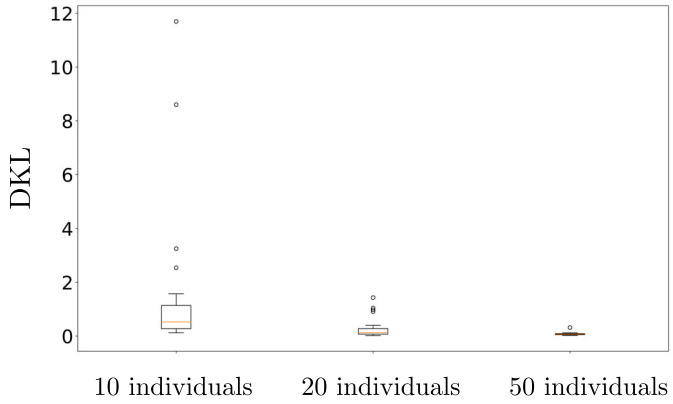


Figure 3.22: Illustration of convergence for mixed-effects calibration of the ODM model. KL divergence between the exact and calibrated distributions for $10, 20$ and 50 individuals.

to estimate the Fisher Information matrix (FIM) which reads as the Hessian matrix of the log-likelihood $\mathcal{H}(-\ln(\mathcal{L}))(\boldsymbol{\Psi})$. In the different experimentations we made, numerical errors were encountered. More precisely, for different configurations of the increment vectors, some of the diagonal terms of $\mathcal{H}(-\ln(\mathcal{L}))(\widehat{\boldsymbol{\Psi}})^{-1}$ were negative, which is inconsistent with the interpretation of $\mathcal{H}(-\ln(\mathcal{L}))(\widehat{\boldsymbol{\Psi}})^{-1}$ as a covariance matrix, and this for 10, 20 and 50 individuals. In fact, it seems that in addition to the error due the approximation of the second order derivative, the inversion of the 9×9 Hessian matrix brings its own amount of error. One way that guarantees a proper estimation of the uncertainty is the bootstrap.

Performing bootstrap requires to choose the re-sampling method. To remind the different options, parametric and non-parametric (or case) bootstrap are available [Lavielle, 2014]. To select the relevant method, it is necessary to remember that in practice, bootstrapped data is a sub-sample of the original data. When the number of original data (here the individuals) is limited, with the non-parametric version of bootstrap, the variety of the bootstrap data can be restrained, artificially reducing the uncertainty of the population parameters. Thus, in our context, the parametric version is preferred to provide relevant estimation of the empirical distribution of $\boldsymbol{\Psi}$, that is thus performed for all 10, 20 and 50 individuals. Still, the main issue of bootstrap is that to reach the convergence of the bootstrap estimator, it requires as many bootstrap samples as possible. Given that a bootstrap sample is synonymous of a negative log-likelihood minimization, this can lead to considerable computational issues. Thus, uncertainty quantification will be only performed for one repetition (the 1^{st} one). Considering as maximum duration of the calculation a week, with 50 individuals, $N_B = 614$ bootstrap samples $\boldsymbol{\Psi}_{\text{boot}}$ can generated. The same calculation took about 3 days for 20 individuals and 1 day for 10 individuals. From the bootstrap samples, we can determine confidence intervals. Remember that the confidence intervals can be determined from the percentiles or using a Gaussian approximation.

Table 3.13: 95 % confidence intervals of the mean parameters for the 1^{st} repetition with 10, 20 and 50 individuals. All parameters are assumed independent (continued with standard-deviations next page).

	E_0 [GPa]	y_{0s} [kPa]	y_{cs} [MPa]	d_c [-]
Confidence interval for 10 individuals on the mean parameters	[175, 199]	[13.7, 26.4]	[1.70, 2.91]	[3.64, 4.95]
Confidence interval for 20 individuals on the mean parameters	[176, 195]	[15.1, 28.4]	[1.95, 2.97]	[3.42, 4.41]
Confidence interval for 50 individuals on the mean parameters	[180, 193]	[19.5, 27.8]	[2.02, 2.44]	[3.53, 4.10]

Table 3.14: 95 % confidence intervals of the standard-deviations parameters for the 1st repetition with 10, 20 and 50 individuals. All parameters are assumed independent.

	E_0 [GPa]	y_{0s} [kPa]	y_{cs} [MPa]	d_c [-]
Confidence interval for 10 individuals on the standard-deviations parameters	[1.66, 21.7]	[0.300, 9.08]	[0.394, 1.15]	[0.429, 1.25]
Confidence interval for 20 individuals on the standard-deviations parameters	[8.17, 20.1]	[0.300, 8.38]	[0.490, 1.04]	[0.606, 1.25]
Confidence interval for 50 individuals on the standard-deviations parameters	[13.0, 22.0]	[3.03, 9.99]	[0.406, 0.687]	[0.750, 1.14]

In Tables 3.13 and 3.14, the 95 % confidence intervals are determined. The lower and upper bounds respectively correspond to the 2.5% and 97.5% quantiles of the empirical distribution of the bootstrap samples. Let's focus first on all parameters except the damage threshold. Clearly, we can see that increasing the number of individuals reduces the uncertainty on the population parameters, either for the means or the standard-deviations. Notice that the reduction of the width of the confidence interval is more significant between 20 and 50 individuals than between 10 and 20 individuals, hinting that adding "only" another 10 individual is not sufficient to reduce much the uncertainty on the estimation $\mathbf{\Psi}$. It is also interesting to remark that the reduction of the uncertainty is more important for the standard-deviation than for the mean parameters. For instance, the ratio of the difference of the upper and lower bounds divided by the MLE estimate of $\mathrm{sd}(d_c)$ is equal to 100% with 10 individuals and drops to 26.3% with 50 individuals while it goes from 34.1% to 15.1% for μ_{d_c}. In fact, this can be explained by remarking that the mean parameters are already estimated with small error with 10 individuals and the 95% confidence interval represents mostly the impact of the lack of specimens. On the contrary, the standard-deviations are not fully converged with only ten individuals. Thus, the 95% confidence intervals account not only for the uncertainty due to the lack of specimens but also for the estimation error, and adding supplementary individuals allows to precise this approximation. The mean of the damage threshold follows the same behavior as for the other parameters. However, we can point out that the uncertainty on the standard-deviation does not reduce with the number of individuals. This behavior is expected considering the small sensitivity of the likelihood function with respect to this model parameter. This is consistent with the fact that the confidence interval for this parameter refers in fact to the interval defined by the bounds of Table 3.2. Observe that the exact set of parameters is always included within the confidence intervals described in Tables 3.13 and 3.14, except for the standard-deviation of the damage evolution celerity. This may be explained by the incomplete coverage of the support empirical bootstrap distributions because of the lack of samples. We can also notice that as the number of individuals increases, the bootstrap samples get centered around the exact set of population error, which is consistent with the reduction of the error on the population parameters (as discussed earlier in this section). However, the confidence interval of Tables 3.13 and 3.14 does not provide indications on the shape of the empirical distribution. An interesting indicator is the coefficient of variation detailed in Table 3.15, which shows that globally, they decrease as the number of individuals increase.

Table 3.15: Coefficient of variation in % of the bootstrapped samples of the population parameters for the 1st repetition with 10, 20 and 50 individuals.

	E_0	y_{0s}	y_{cs}	d_c
COV of the mean bootstrap samples for 10 individuals [%]	3.42	16.5	12.7	7.49
COV of the mean bootstrap samples for 20 individuals [%]	2.72	14.9	9.76	6.66
COV of the mean bootstrap samples for 50 individuals [%]	1.73	8.83	4.93	4.25
COV of the standard-deviations bootstrap samples for 10 individuals [%]	36.8	94.1	25.9	25.7
COV of the standard-deviations bootstrap samples for 20 individuals [%]	22.4	104	19.4	18.2
COV of the standard-deviations bootstrap samples for 50 individuals [%]	13.3	23.9	12.9	4.17

In fact, the indicators show that there is a reduction of the uncertainty such as a concentration around the mean value of the empirical distributions (the COVs proposed in Table 3.15). Figure 3.23, which displays some marginals of the empirical distributions of the bootstrap samples for the first repetition with 10, 20 and 50 individuals summarizes these observations (the others and the correlations are available in Figure 3.24).

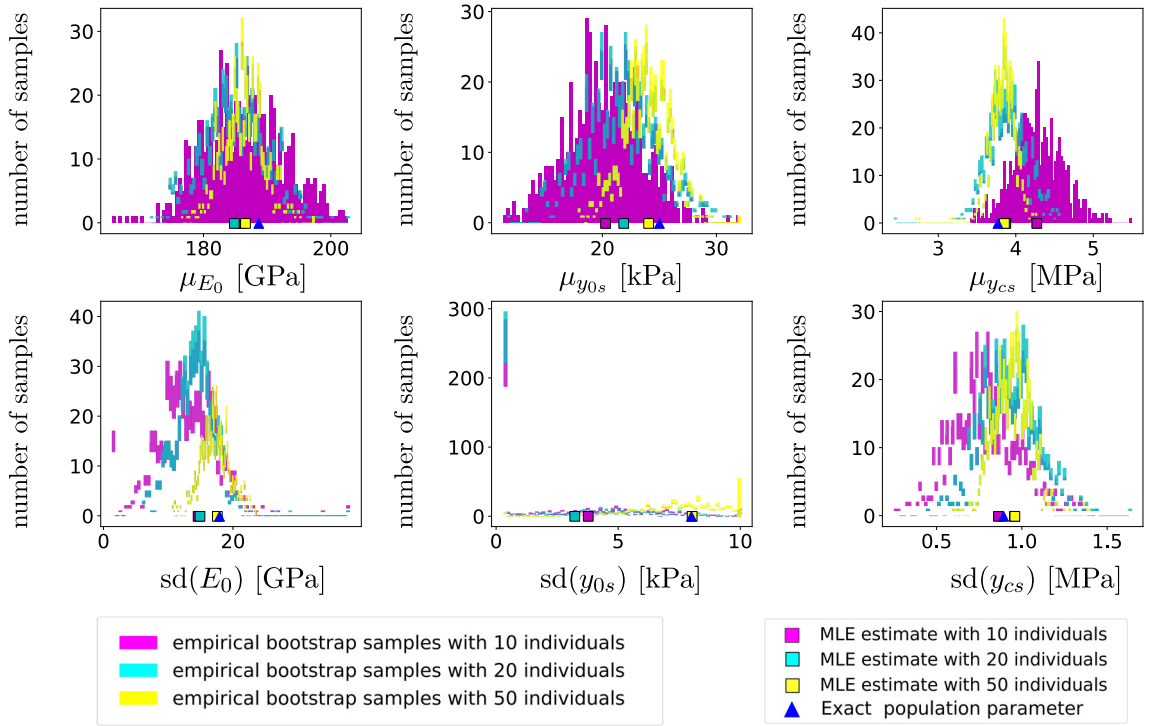


Figure 3.23: For the first repetition, empirical distributions of the bootstrap samples for μ_{E_0} (top left), $\mu_{y_{0s}}$ (top center), $\mu_{y_{cs}}$ (top right), sd(E_0) (bottom left), sd(y_{0s}) (bottom center) and sd(y_{cs}) (bottom right) for 10, 20 and 50 individuals. Along the bootstrap samples, exact values and MLE estimates.

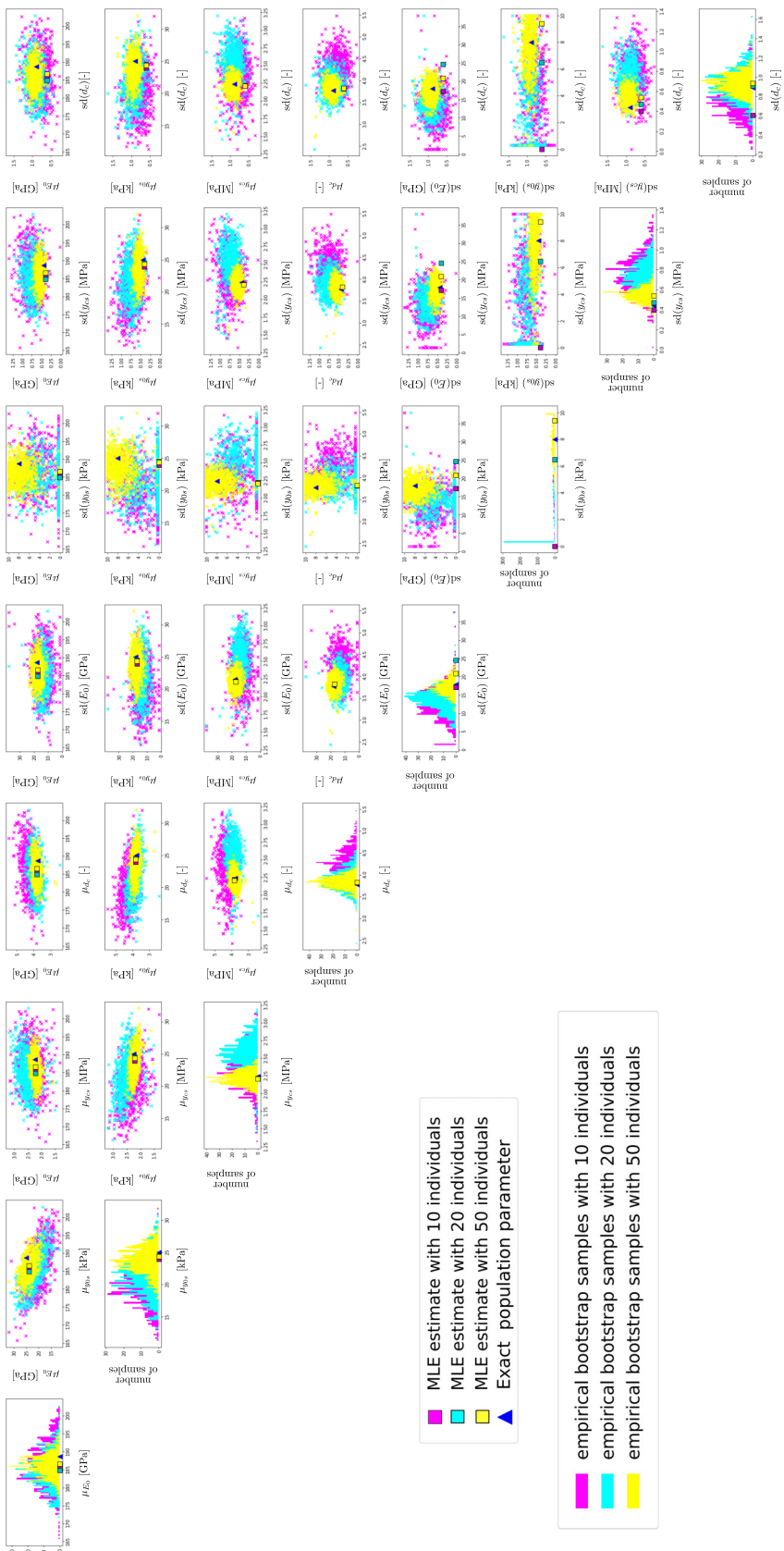


Figure 3.24: For the first repetition, empirical distributions of the bootstrap samples for all population parameters with 10, 20 and 50 individuals. Along the bootstrap samples, exact values and MLE estimates.

From both of these indicators, it follows that, as expected, there is a clear reduction of both the error and uncertainty decrease with the number of individuals but can remain significant with 50 individuals. Still, such an increase is synonymous with an increase of the duration of the calculations (because there are more individual parameters to estimate). This point is evaluated further below.

3.4.3.2 Evaluating the impact on the computational costs

Aside from the extra financial costs (the price of the material, the use of the testing equipment, technician hours, *etc.*) implied by the increase of the number of individuals that comes from the new experiments to carry out, one of the other side effects is the increase of the computational costs because there are more individual calibrations to carry out. This is a major concern as the individual calibrations refer to the costly phase of the likelihood computation. To examine the evolution of the computational costs, it is possible to take a look to Figure 3.25 for different number of individuals.

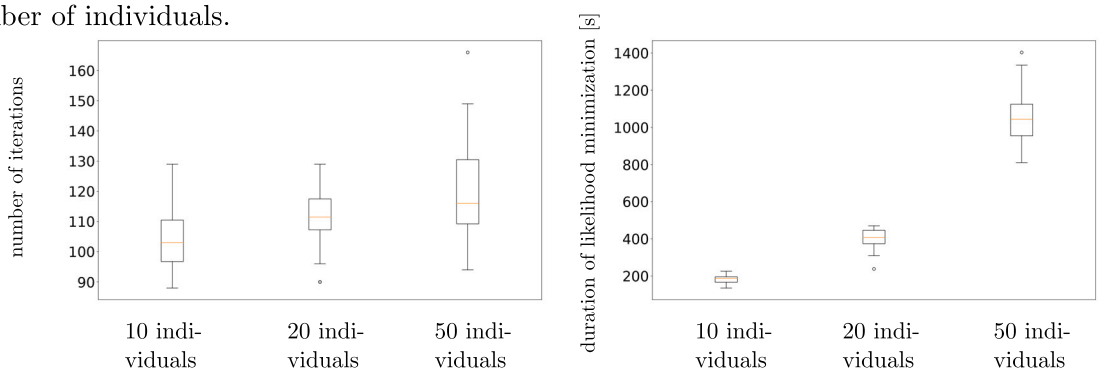


Figure 3.25: Computational costs implied by mixed-effects methods. On the left, number of iterations for the likelihood minimization with different number of individuals. On the right, duration of the likelihood minimization for different number of individuals.

Figure 3.25 shows that the number of likelihood evaluations is roughly the same for all number of individuals, even if it can be noticed that the number of iterations of the minimization slightly increases (and so does the number of likelihood evaluations) when the number of individual increases (the median number of the likelihood increases by 9%). Yet, the duration of the optimization inflates when supplementary individuals are added (multiplication by 6 from 10 to 50 individuals), but this phenomenon remains limited to linear trend which can be considered as acceptable. To be more precise, the complexity of the log-likelihood calculation is linear with respect to the number of individuals, because the number of auxiliary minimizations necessary to compute the approximate log-likelihood is equal to the number of individuals. Thus, the number of specimens required should be estimated by striking a balance between the precision requirements and the protocol expenses which consists of, both, the number of experiments to be carried out and the computational costs necessary to achieve the calibration. Notice that, between those two, the effective limiting factor is more likely the size of the experimental campaign.

This section has illustrated the influence of the number of individuals. It showed that as we could expect, adding individuals both reduces the estimation error of the population parameters but also their uncertainty. Yet, all this study has only been performed with a simplistic modeling (*i.e.*, multivariate Gaussian distribution with independent copula), which does not take into account the potential of the couplings between the model parameters. This topic is studied in the next section.

3.4.4 Calibrating with dependency structure

Up to now, the chosen modeling for the distribution of the calibration parameters did not involve dependencies between the model parameters. The objective is to present first the notion

of dependency, before describing the numerical implementation. Then, the estimation on the population parameters are checked with of course a focus on the hyper-parameters that characterize the dependency structure which refers here to the correlations. Figure 3.26 sums up the different steps of the analysis.

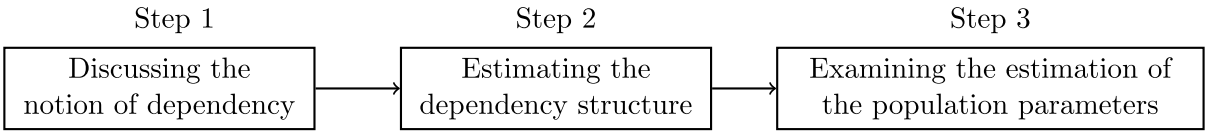


Figure 3.26: Flowchart of the steps for the evaluation of a dependency structure with mixed-effects.

3.4.4.1 Introducing the notion of statistical dependency

Let us first explain the different notions of dependency. Indeed, most behavior laws imply dependencies between the parameters. Yet, the concept of parameter interaction or coupling can be understood in different ways and deserves some clarification.
A first kind of coupling comes from the model formulation. This can be illustrated with the ODM model: when damage reaches saturation, $d_s(t) \approx d_c$ (Eq. (1.24c)), the effective modulus reads as $E^{\text{eff}}(d_s(t)) \approx \frac{E_0}{1+d_c}$ (Eq. (1.23)) and the model simplifies as:

$$\sigma(t) \simeq \frac{E_0}{1+d_c}\varepsilon(t) \ . \tag{3.14}$$

Thus, an increase of the damage saturation can compensate an increase of the Young's modulus and vice versa. Here, the Young's modulus and the damage saturation are positively correlated through the model. These interactions through the model formulation can be thought of as functional dependencies. Such interaction is taken into account in the determination of the individual parameters *via* the first term of function $g_i(\cdot)$ (defined in Eq. (3.5)) that describes how well the model output $\mathcal{F}(\mathbf{t}_i, \boldsymbol{\mu} + \mathbf{b}_i)$ matches with the experiment, that is to say $-\ln(f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega))$.
With the mixed-effects models, a second category of dependencies emerges when a dependency structure is added to $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$. These dependencies will be called statistical dependencies. Given a constitutive law with parameters $\boldsymbol{\Theta}$, some terms in the higher level parameters $\boldsymbol{\Pi}$ represent statistical dependencies between the $\boldsymbol{\Theta}$s. Contrary to the first dependencies, these interactions are informative about the different phenomena that interact with each other (elasticity, damage, *etc.*). For instance, in general, the parameters that characterize the fiber behavior can be considered as physically (and so statistically) independent from those that are associated to the matrix behavior. However, the parameters describing the shear behavior, which combine both fiber and matrix behaviors will be dependent on the parameters that describe the matrix and fiber behaviors (and conversely). Still, these physical dependencies are read through the model parameters, and therefore, the interpretations should be taken with caution. In the ODM constitutive law, an increase in the Young's modulus E_0 is physically associated to a faster development of damage (a smaller y_{cs}). These interactions are taken into account in the determination of the individual parameters *via* the second term in function $g_i(\cdot)$, $\|\boldsymbol{\Delta}_i \mathbf{b}_i\|_2^2$, that weights how well individual parameters correspond to the higher level distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$. They do not refer to spatial correlations that express within the same composite sheet studied in [Sriramula and Chryssanthopoulos, 2009] neglected in this first approach considering a whole composite sheet as a single realization of the material properties. These correlations are of prime importance as neglecting them can lead to erroneous computations of design quantities [Smarslok et al., 2012].

3.4.4.2 Numerical estimation of the statistical dependency structure

In this section, the objective is to characterize material variability and the statistical dependencies between the different phenomena driving the parameters values. For a Gaussian random

vector of parameters, the statistical dependency is determined by the covariance matrix (here $\mathbf{\Sigma}$). If all components are independent, $\mathbf{\Sigma}$ is a diagonal matrix, otherwise off-diagonal terms describe the statistical dependencies. To explore different correlation scenarios, we choose a weak anti-correlation ($\rho(E_0, y_{cs}) = -0.4$) and a strong correlation ($\rho(d_c, y_{cs}) = 0.8$) (where ρ stands for the correlation parameter). These values are arbitrarily chosen as, to the best of our knowledge, there is no prior information on their values in the literature for this type of material. All the other parameters of the exact distribution (means, standard-deviations) remain the same as in the previous independent case. Given the modeling choices, $\mathbf{\Sigma}$ is parameterized by 4 variances and 2 covariances terms. It can be detailed into:

$$\mathbf{\Sigma} = \begin{array}{c} \\ E_0 \\ y_{0s} \\ y_{cs} \\ d_c \end{array} \begin{array}{c} \begin{array}{cccc} E_0 & y_{0s} & y_{cs} & d_c \end{array} \\ \begin{pmatrix} \mathbb{V}(E_0) & * & * & * \\ 0 & \mathbb{V}(y_{0s}) & * & * \\ Cov(E_0, y_{cs}) & 0 & \mathbb{V}(y_{cs}) & * \\ 0 & 0 & Cov(y_{cs}, d_c) & \mathbb{V}(d_c) \end{pmatrix} \end{array}, \tag{3.15}$$

where the symbol $*$ stands for the symmetric coefficient of the covariance matrix. The calibration of the covariance matrix is performed using a Cholesky decomposition to avoid numerical constraints between the correlation parameters (*i.e.*, positive-definiteness of $\mathbf{\Sigma}$) [Pinheiro and Bates, 1996]. The Cholesky decomposition of $\mathbf{\Sigma}$, noted $\mathbf{L}$, reads as:

$$\mathbf{L} = \begin{pmatrix} L_{11} & 0 & 0 & 0 \\ 0 & L_{22} & 0 & 0 \\ L_{31} & 0 & L_{33} & 0 \\ 0 & 0 & L_{43} & L_{44} \end{pmatrix} . \tag{3.16}$$

Except L_{11} and L_{22} which correspond to the standard-deviations of E_0 and y_{0s}, the other terms have no direct interpretation as covariances or standard-deviations. It makes difficult to determine bounds on these coefficients to define the search space associated to the minimization of the negative log-likelihood. The Cholesky decomposition may be also expressed with a spherical decomposition [Pinheiro and Bates, 1996]. This consists in decomposing the rows in a generalized coordinate system as follows:

$$\mathbf{L} = \mathbf{SO} , \tag{3.17}$$

with $\mathbf{S} \coloneqq \sqrt{\text{diag}(\mathbf{\Sigma})}$. The diagonal terms of this matrix refer to the standard-deviation of the model parameters. The general form of matrix $\mathbf{O}$ is:

$$\mathbf{O} \coloneqq \begin{pmatrix} \cos(\nu_{11}) & \sin(\nu_{11})\cos(\nu_{12}) & \sin(\nu_{11})\sin(\nu_{12})\cos(\nu_{13}) & \sin(\nu_{11})\sin(\nu_{12})\sin(\nu_{13}) \\ \cos(\nu_{21}) & \sin(\nu_{21})\cos(\nu_{22}) & \sin(\nu_{21})\sin(\nu_{22})\cos(\nu_{23}) & \sin(\nu_{21})\sin(\nu_{22})\sin(\nu_{23}) \\ \cos(\nu_{31}) & \sin(\nu_{31})\cos(\nu_{32}) & \sin(\nu_{31})\sin(\nu_{32})\cos(\nu_{33}) & \sin(\nu_{31})\sin(\nu_{32})\sin(\nu_{33}) \\ \cos(\nu_{41}) & \sin(\nu_{41})\cos(\nu_{42}) & \sin(\nu_{41})\sin(\nu_{42})\cos(\nu_{43}) & \sin(\nu_{41})\sin(\nu_{42})\sin(\nu_{43}) \end{pmatrix} . \tag{3.18}$$

Given the form described in Eq. (3.16), $\nu_{11}, \nu_{22}, \nu_{33}$ equal to 0 and $\nu_{21}, \nu_{32}, \nu_{41}, \nu_{42}$ equal to $\frac{\pi}{2}$. The covariance matrix is parameterized by ν_{31} and ν_{43} and all others terms do not influence the result and are thus not taken into consideration. Note that the correlation matrix $\mathbf{R}$ can be directly computed as $\mathbf{R} = \mathbf{OO}^\top$. In total, there are $\ell = 4$ (standard-deviations) +2 (angles) +4 (means) +1 (noise) = 11 parameters to calibrate. Furthermore, whichever method is chosen to parameterize the covariance matrix, the number of population parameters remains the same. Here, the angles ν_{ij} are all bounded between 0 and π, the other parameters referring to the standard-deviations that are easy to bound. In addition, there is direct a link between the correlations and the angles. More precisely, the correlations ρ_{i1} in the first column can be directly expressed from the angles ν_{i1}, which gives here $\rho(E_0, y_{cs}) = \rho_{31} = \cos(\nu_{31})$. The other correlation reads as $\rho(d_c, y_{cs}) = \rho_{43} = \sin(\nu_{31})\cos(\nu_{43})$. Issues of positive-definiteness can

still occur when the angles hit the mathematical bounds (0 or π) and can cause the loss of the positive-definite property, but this can be prevented by reducing the bounds, for instance taking 10^{-5} instead of 0 and $(1-10^{-5})\pi$ instead of π. In the following and for the remainder of this work, we implement the spherical parameterization.

Before proceeding to the analysis, the behavior of the minimization is briefly discussed. The estimation of the marginal parameters is first checked. To ensure that the estimation of the marginals is not deteriorated, it is possible to check to what extent their estimation is impacted and especially if it adds variability on the marginal parameters. As in the previous sections, the relative difference between the best estimation of the marginal parameters (in terms of likelihood) $\widehat{\mathbf{\Pi}}_{\text{marg,best}}$ and the three others $\widehat{\mathbf{\Pi}}_{\text{marg}}$, defined as $\delta(\widehat{\mathbf{\Pi}}_{\text{marg}}, \widehat{\mathbf{\Pi}}_{\text{marg,best}}) := 100 \times \frac{|\widehat{\mathbf{\Pi}}_{\text{marg}} - \widehat{\mathbf{\Pi}}_{\text{marg,best}}|}{\widehat{\mathbf{\Pi}}_{\text{marg,best}}}$, are computed and averaged over the remaining calibrations for the 9^{th} repetition in Table 3.16 considering 10, 20 and 50 individuals.

Table 3.16: Averaged relative difference δ in % of the mean and standard-deviation parameters for different initializations for the 9^{th} repetition of the calibration process. 10 individuals are considered. The random seed remains the same for all calibrations.

Population Parameter		μ_{E_0}	$\mu_{y_{0s}}$	$\mu_{y_{cs}}$	μ_{d_c}
Averaged relative difference [%]		0.0205	0.0332	0.0287	0.0373
Population parameter	sd(E_0)	sd(y_{0s})	sd(y_{cs})	sd(d_c)	ω
Averaged relative difference [%]	0.167	0.0186	0.0927	0.0883	0.0449

Tables 3.16 shows that for the 4 initializations, the identified marginal parameters are almost the same (the relative difference reaches at most 0.1%), which shows that there is no specific deterioration of their estimation. In particular, it does not add uncertainty to the estimation of the parameters.

It is possible now to study the calibration of the correlations. To assess their estimation, the candidate correlations selected throughout the minimization for 10, 20 and 50 individuals are examined in Figure 3.27.

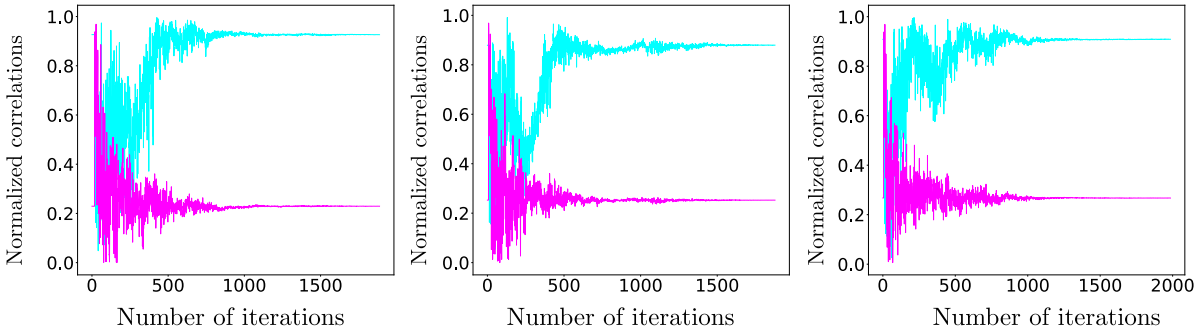


Figure 3.27: Candidate normalized correlations through the optimization for the 9^{th} repetition with 10 (left), 20 (center) and 50 (right) individuals. The cyan line refers to the $\rho(E_0, y_{cs})$ and the magenta line to the $\rho(d_c, y_{cs})$.

Figure 3.27 shows that for all the number of individuals, the correlation parameters demonstrate a proper convergence behavior towards their final estimation, suggesting that their estimation do not cause specific numerical issues. With a dependency structure, the calibration with correlations demands more iterations to converge, on average about 33% iterations for all number of individuals. This increase is not surprising: there are more parameters to calibrate, requiring a more complex exploration of the parameters space. We can also check to what extent the estimation of the correlation coefficients depends on the starting points. This can be evaluated by computing the averaged relative difference δ between the correlations for the best calibration $\widehat{\rho}_{\text{best}}$ (in terms of likelihood) and those for the three others $\widehat{\rho}$, defined as $\delta(\widehat{\rho}_{\text{best}}, \widehat{\rho}) := 100 \times \frac{|\widehat{\rho} - \widehat{\rho}_{\text{best}}|}{\widehat{\rho}_{\text{best}}}$,

then averaged over the three remaining calibrations and reported in Table 3.17 for the 9^{th} repetitions of the calibration process and different number of individuals.

Table 3.17: Averaged relative difference δ in % of the correlations parameters for different initializations for the 9^{th} repetition with 10, 20 and 50 individuals. The random seed remains the same for all calibrations.

	$\rho(E_0, y_{cs})$	$\rho(d_c, y_{cs})$
Averaged relative difference [%] with 10 individuals	1.51	0.0572
Averaged relative difference [%] with 20 individuals	0.179	0.0300
Averaged relative difference [%] with 50 individuals	0.178	0.0740

Table 3.17 shows that, even with few individuals, the estimation of the correlation parameters does not appear to depend much on the choice of initialization points, but to a lesser extent than the marginals parameters. Because of the limited number of repetitions and thus the higher numerical noise, the levels of difference are the same for all numbers of individuals. In the most pessimistic case, the averaged relative difference remains below 2%, for the smallest correlation in absolute value. This is not necessarily surprising as a small correlation value does not coerce much the individuals to follow a given pattern, hinting for a limited sensitivity of the likelihood function with respect to this value. Nevertheless, this phenomenon remains limited and does not prevent a reliable estimation of the correlations, allowing to proceed to the analysis of the results in the next section.

3.4.4.3 Estimation of the population parameters

The testing procedure implemented in the previous sections to examine the calibration of the population parameters is applied here. Before analyzing the calibration of the correlations, a brief analysis of the estimation of the marginal and individual parameters is carried out. The corresponding results are reported in Tables 3.18 and 3.19 .

Table 3.18: Averaged error $\mathcal{E}_{\text{mean}}(\widehat{\mathbf{\Pi}})$ in % for the marginal parameters (means and standard-deviations) over 20 repetitions of the calibration process with samples of 10, 20 and 50 specimens each. The coefficients of variation COV in % are indicated between brackets. Calibration results with dependency structure.

	E_0	y_{0s}	y_{cs}	d_c
Averaged error on mean parameters for 10 individuals [%]	3.09 (90.1)	7.61 (78.5)	6.17 (71.2)	5.97 (70.7)
Averaged error on mean parameters for 20 individuals [%]	1.65 (73.0)	6.25 (74.2)	4.31 (82.1)	4.18 (90.2)
Averaged error on mean parameters for 50 individuals [%]	1.25 (67.4)	4.24 (74.4)	2.83 (63.8)	2.83 (84.4)
Averaged error on standard-deviations parameters for 10 individuals [%]	21.5 (83.0)	45.4 (70.6)	23.8 (82.0)	17.5 (82.7)
Averaged error on standard-deviations parameters for 20 individuals [%]	17.0 (78.9)	41.2 (85.9)	15.7 (79.5)	15.4 (65.9)
Averaged error on standard-deviations parameters for 50 individuals [%]	12.0 (74.4)	24.8 (103)	12.7 (63.5)	8.33 (87.5)

Table 3.19: Comparison of the relative errors $\mathbf{e}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_{50})$ in % on the individual parameters averaged over 20 repetitions of the calibration process with of 10, 20 and 50 specimens individuals. The corresponding coefficients of variation in % are indicated in brackets. Calibration results with dependency structure.

	E_0	y_{0s}	y_{cs}	d_c
Averaged relative error on the individual parameters with 10 individuals [%]	3.05(28.7)	22.6(44.5)	7.86(23.4)	4.15(24.4)
Averaged relative error on the individual parameters with 20 individuals [%]	3.05(21.2)	21.7(29.1)	7.19(18.3)	4.03(14.7)
Averaged relative error on the individual parameters with 50 individuals [%]	2.83(17.2)	23.4(43.1)	6.76(16.4)	3.82(17.2)

The major conclusion from Tables 3.18 and 3.19 can be summed up as follows: in our context (that is to say given the modeling and the parameter values), adding a dependency structure to $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ deteriorates neither the estimation of the marginals nor the estimation of the individuals parameters. The standard-deviations parameters may have a small increase in the error when no dependency structure is assumed. Indeed, to decrease the likelihood, the optimization can now rely on two extra population parameters when no dependency structure is assumed. In fact, adding two extra parameters within $\boldsymbol{\Psi}$ does not seem to alter the minimization process and consequently the determination of both the means and standard-deviations. We can now proceed to the analysis of the calibrated correlations reported in Table 3.20.

Table 3.20: For 20 repetitions with 10, 20 and 50 specimens, averaged calibrated correlations $\widehat{\boldsymbol{\Pi}}_{\text{mean}}$ and associate averaged relative error $\mathcal{E}_{\text{mean}}(\widehat{\boldsymbol{\Pi}})$ in %. The corresponding coefficients of variation COV in % are indicated between brackets.

	$\rho(E_0, y_{cs})$	$\rho(d_c, y_{cs})$
Exact values	-0.4	0.8
Calibrated correlations with 10 individuals	-0.335(83.4)	0.811(17.0)
Averaged errors on the correlations with 10 individuals [%]	58.6 (70.9)	14.3 (67.2)
Calibrated correlations with 20 individuals	-0.361 (47.0)	0.808 (10.0)
Averaged errors on the correlations with 20 individuals [%]	39.1 (49.0)	8.91 (64.0)
Calibrated correlations with 50 individuals	-0.370 (41.0)	0.798 (8.00)
Averaged errors on the correlations with 50 individuals [%]	30.7 (76.6)	6.65 (66.4)

With few individuals (10), over 20 repetitions of the calibration process, the estimation of both correlations turns out to be consistent with the exact values. Note that the COVs of the correlations of $\rho(E_0, y_{cs})$ in Table 3.20 is positive, even if the target value is negative (-0.4) (we take the standard-deviation as the opposite square-root of the empirical variance). It shows that the average relative error of the correlation parameters is larger between $\rho(E_0, y_{cs})$ than $\rho(d_c, y_{cs})$, and this for all repetitions (the relative error on $\rho(E_0, y_{cs})$ even reaches 100% for 2 repetitions). In fact, this comes from the fact that $\rho(d_c, y_{cs})$ is the highest correlation, which means that among the two correlations, it refers to the one that coerces the most the individual parameters and thus, the likelihood is more sensitive to it compared to the other one ($\rho(E_0, y_{cs})$), explaining why it is better calibrated than the other (*e.g.*, between E_0 and y_{cs}) regardless of the number of specimens. As the number of individuals increases, the calibrated correlations tend to the exact value. Nevertheless, even if the correlation $\rho(E_0, y_{cs})$ is better

identified with 50 individuals than with 10, an even larger number of individuals seems necessary to correctly calibrate this statistical dependency. The convergence of the estimation of the correlation between E_0 and y_{cs} for the 9$^{\text{th}}$ repetition is illustrated in Figure 3.28.

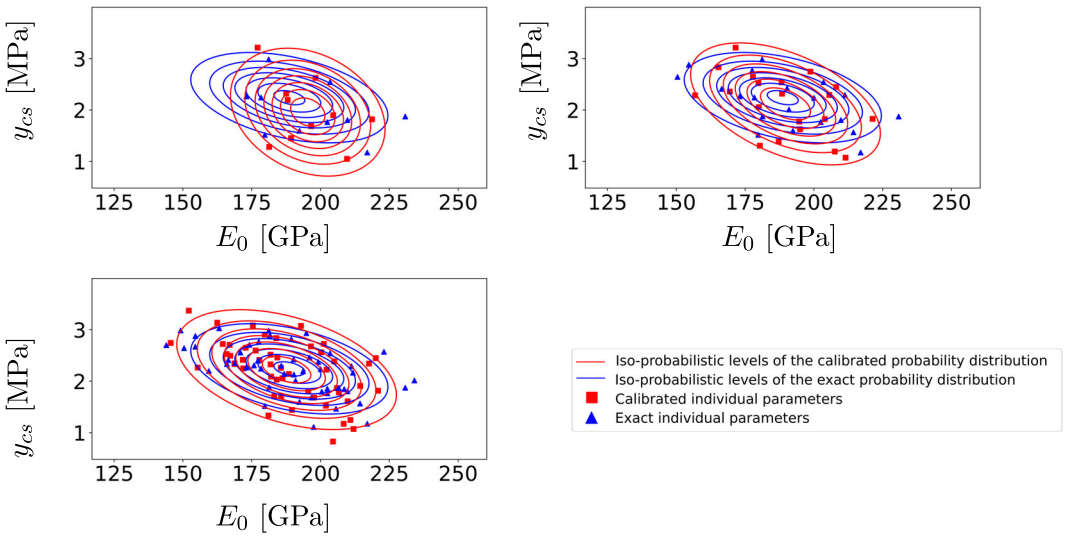


Figure 3.28: Illustration of convergence for mixed-effects calibration of the ODM model in parameters space with the exact and calibrated copulas between E_0 and y_{cs} estimated with 10 (a), 20 (b) 50 individuals (c) for the 9$^{\text{th}}$ repetition. Individual parameters are also plotted.

These results clearly show that the number of available specimens is a key issue in population-based approaches. Given the estimation errors affecting the method, we think that statistical dependencies should be used carefully, especially if the number of available specimens is small and the interactions are weak. Therefore, it is questionable whether with few individuals taking into account the correlation structure represents an improvement or not. To test this point, we calibrate the ODM model using mixed-effects with the same data without the consideration of the dependency structure (so we have 9 parameters to calibrate as the in previous sections) for 10, 20 and 50 individuals. Next, we can investigate whether considering a statistical dependency provides better estimates of the model parameters distribution or not with respect to the number of individuals. This can be done with the help of the KL divergence that measures how close the calibrated distribution is to the exact one, computed for 10, 20 and 50 individuals for both calibration scenarios in Figure 3.29.

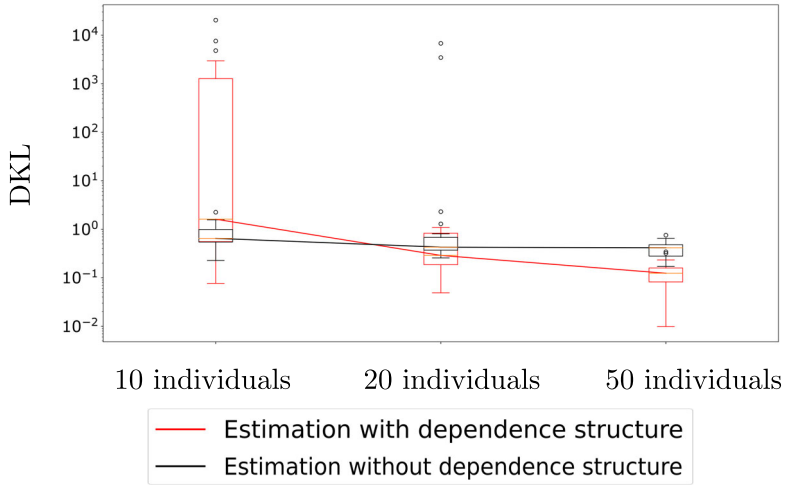


Figure 3.29: Comparison of the calibration considering including (red) or not (black) statistical dependencies for 10, 20 and 50 individuals based on the DKLs. The lines refer to the median trajectories.

On the left of Figure 3.29, we can see that with 10 individuals, the estimation of the model parameters distribution is better with an inexact assumption (*i.e.* without a dependency modeling) rather than for the exact assumption (*i.e.* including a dependency structure). On the contrary, with 50 individuals and 20 to a lesser extent, the estimated of $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ is better if it based on an exact assumption. It is interesting to notice that, without statistical dependency, the KL divergence still decreases with the number of individuals. However, it is much slower than what can be observed on the left of Figure 3.22. Especially, without dependency structure, the KL divergence do not seem to converge to 0 but seems to be limited by a lower bound, that may refer to the KL divergence between the exact distribution and the distribution with the same marginal parameters but without statistical dependencies. This lower bound could be interpreted as a sort of "model bias", the word "model" referring here to $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$. On the contrary, the convergence of the estimated distribution to the exact distribution exhibits a similar behavior as the one observed in Figure 3.22. This comes from the fact that the correlations estimated with 10 individuals show significant error (58.6% on average for $\rho(E_0, y_{cs})$) while this error drops when adding more individuals (30.7% on average for $\rho(E_0, y_{cs})$ with 50 individuals), allowing to get closer the exact distribution, explaining the observed behavior.

This section presented the tools to include in the modeling a Gaussian dependency structure, characterized by the correlation matrix, for Gaussian distributed parameters. The methods to estimate these correlations (in particular the spherical decomposition) have been introduced. The estimations of correlations have been implemented and the different results studied. It has shown that the inclusion of statistical dependencies does not alter the calibration of the marginals and of the individual parameters or the convergence of the minimization algorithm. We have seen that mixed-effects methods for model calibration have been shown their ability to estimate the correlations, either weak and strong depending on the number of available specimens. However, we could notice that the averaged errors on the weak correlations is much larger than on the strong correlations, even if they decrease with the number of individuals. Finally, we illustrated the importance of the number of individuals that highlights to be a key feature either in the modeling or on the estimation of a dependency structure. For the rest of this section, all numerical simulations will be performed with 50 individuals to ensure statistical consistency of the estimators.

3.4.5 Reconsidering different assumptions on model parameters distribution

In real cases, the exact model parameter distribution is unknown and limiting ourselves to a Gaussian multivariate distribution may seem too restrictive. Though this assumption is widely used because of both modeling and computational motives, it has numerous drawbacks such as a perfect symmetry. In addition, the theoretical background of mixed-effects is not limited to the choice of Gaussian distributions. This section is devoted to the study of mixed-effects calibration considering different assumptions on the model parameters distribution and on the dependency structures. The consequences on the likelihood estimation are discussed. Then, the data used in the previous section is processed with mixed-effects to check their ability to estimate the population parameters for different types of distribution and to see whether assumptions can be rejected or not.
To test the influence of hypotheses, two different assumptions are tested for the marginal of the parameter to which the model output is the most sensitive (the Young's modulus). Most of the time, the variability of material properties is modeled by a Gaussian distribution [Bright et al., 2011b, Niyigena et al., 2016]. Thus, we choose for the other marginal a distribution whose shape is similar to that of the Gaussian but is non-symmetric : the Gumbel distribution. Furthermore, it complies with the assumptions of the Laplace approximation. The reason why the Gumbel distribution is interesting that it is non-symmetric. Even if the natural parameters (β, γ) of the latter distribution do not refer to mean and standard-deviations, they can be easily

related to them by $\beta = \text{sd}(E_0)\sqrt{\frac{6}{\pi}}$ and $\gamma = \mu_{E_0} - \gamma_e \text{sd}(E_0)\sqrt{\frac{6}{\pi}}$ with γ_e the Euler-Mascheroni constant approximately equal to 0.577. Thus, we can achieve the calibration using mean and standard-deviations, simplifying the parameterization of the minimization problem and easing the comparison of the results. The copula remains either Gaussian or independent for the sake of simplicity. Indeed, with few individuals, estimating a Gaussian copula is already tricky and may be even more harder for other copula models. In addition, note that these other model may also not always satisfy the regularity conditions of the Laplace approximation. Thus, the number of parameters to estimate remains the same: 9 without correlations and 11 otherwise that still refer to 4 means, 4 standard-deviations, 1 variance of the error term and 2 eventual extra correlations. The different calibration scenarios are summarized in Table 3.21, where the number of parameters includes, aside from the marginals and copula parameters, the standard-deviation of the residual model.

Table 3.21: Different assumptions for the calibration of the ODM model with mixed-effects alongside their associated colors.

Scenario	Marginals	Copula	Number of parameters to estimate
$\mathcal{S}_1$	Gaussian	Independent	9
$\mathcal{S}_2$	All Gaussian except a Gumbel marginal for E_0	Independent	9
$\mathcal{S}_3$	Gaussian	Gaussian except for y_{0s}	11
$\mathcal{S}_4$	All Gaussian except a Gumbel marginal for E_0	Gaussian except for y_{0s}	11

The first step here intends to show how in practice, other assumptions than Normal can be implemented and how to carry out the likelihood calculations. Then, the calibration from the different scenarios are compared from different points of view that refer to the estimation of the population parameters and individual parameters. The final stage consists in propagating the calibrated distribution in model space to compare the different estimations of material variability. The different steps of the analysis are reported in Figure 3.30.

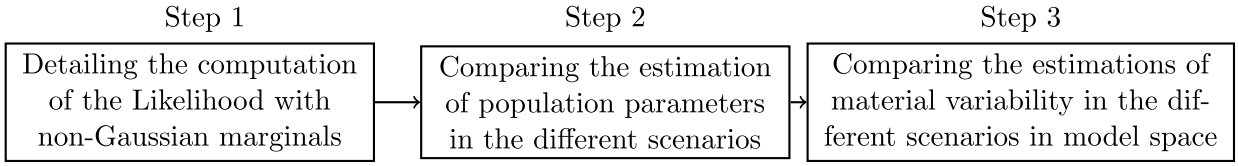


Figure 3.30: Flowchart of the different steps for the evaluation of the consequences of the calibration with non-Gaussian on the mixed-effects calibration results.

3.4.5.1 Estimation of the likelihood

Remember that the Laplace approximation for the log-likelihood is expressed as:

$$\begin{aligned} -\ln(\mathcal{L}(\boldsymbol{\Psi})) &= -\sum_{i=1}^{n} \ln\left(\int_{\mathbb{R}^d} f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) \mathrm{d}\boldsymbol{\theta}_i\right) \\ &\approx \sum_{i=1}^{n} \left(\frac{1}{2}\ln(|\mathcal{H}(g_i)(\widehat{\boldsymbol{\theta}}_i)|) + g_i(\mathbf{y}_i, \widehat{\boldsymbol{\theta}}_i, \boldsymbol{\Psi})\right) - \frac{d}{2}\ln(2\pi) \ , \end{aligned} \tag{3.19}$$

with:

$$g_i(\mathbf{y}_i, \boldsymbol{\theta}_i, \boldsymbol{\Psi}) \coloneqq -\ln(f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})) \ , \tag{3.20}$$

where $f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega)$ refers to the density of the data given the individual and population parameters, $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})$ to the model parameters distribution, d is the dimension and n the number of specimens. Its Hessian matrix can be expanded into:

$$\mathcal{H}(g_i)(\widehat{\boldsymbol{\theta}}_i) = -\mathcal{H}(\ln(f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega)))(\widehat{\boldsymbol{\theta}}_i) - \mathcal{H}(\ln(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})))(\widehat{\boldsymbol{\theta}}_i) \ . \tag{3.21}$$

Here, $f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega)$ does not change and remains proportional to $\exp\left(-\frac{\|\mathbf{y}_i - \mathcal{F}(\mathbf{t}_i, \boldsymbol{\theta}_i)\|_2^2}{2\omega^2}\right)$. Thus, the change in the calculation occurs in the expression of $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})$ and for its Hessian matrix. Remember that any distribution can be decomposed (uniquely) between marginals and the Gaussian copula following Sklar's theorem [Sklar, 1959] as:

$$f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) = c(F_1(\theta_{i1}|\boldsymbol{\Pi}_1), \dots, F_d(\theta_{id}|\boldsymbol{\Pi}_d)|\boldsymbol{\Pi}_{\text{cop}}) \times \prod_{l=1}^{d} f_l(\theta_{il}|\boldsymbol{\Pi}_l) \ , \tag{3.22}$$

with $\boldsymbol{\Pi}_{\text{cop}}$ the copulas parameters, $\boldsymbol{\Pi}_l$ the parameters of the l^{th} marginal with probability density function (PDF) $f_l(\cdot)$ and cumulative density function (CDF) $F_l(\cdot)$ the marginal of the l^{th} coordinate of the i^{th} individual parameter $\theta_{il}, (i,l) \in [\![1,n]\!] \times [\![1,d]\!]$ and $c(\cdot)$, the density function of a copula. Here, the density function of the Gaussian copula is:

$$c(F_1(\theta_{i1}|\boldsymbol{\Pi}_1), \dots, F_d(\theta_{id}|\boldsymbol{\Pi}_d)|\boldsymbol{\Pi}_{\text{cop}}) \coloneqq \frac{1}{\sqrt{|\mathbf{R}|}} \exp\left(-\frac{1}{2}\begin{pmatrix}\Phi^{-1}(F_1(\theta_{i1}|\boldsymbol{\Pi}_l))\\ \dots \\ \Phi^{-1}(F_d(\theta_{id}|\boldsymbol{\Pi}_d))\end{pmatrix}^{\top} \mathbf{U} \begin{pmatrix}\Phi^{-1}(F_1(\theta_{i1}|\boldsymbol{\Pi}_l)))\\ \dots \\ \Phi^{-1}(F_d(\theta_{id}|\boldsymbol{\Pi}_d))\end{pmatrix}\right) \ , \tag{3.23}$$

with $\Phi^{-1}(\cdot)$ the inverse CDF of univariate Gaussian distribution with a 0 mean and a unit variance, $\mathbf{U} \coloneqq \mathbf{R}^{-1} - \mathbf{I}_d$ of coefficients $(u_{lm})_{1 \leqslant l,m \leqslant d}$ and $\mathbf{I}_d$ the identity matrix of $\mathcal{M}_d(\mathbb{R})$. The formula of Eq. (3.23) is semi-analytical as $\Phi^{-1}(\cdot)$ cannot be expressed in closed-form. The evaluation of $\Phi^{-1}(\cdot)$ is precise enough with usual softwares but its accuracy remains an important factor as it should enable to explore the tails of the $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$. Indeed, the Laplace approximation is based on a linearization around the individual parameters that can be located far from the mode value, especially for parameters that have a significant variance. Note that, with no correlation, $\mathbf{R}$ is reduced to the identity matrix and the copula density $c(\cdot)$ to a function with a constant output equal to 1.
Another advantage of the Gaussian copula is that it enables an accurate estimation of the Hessian matrix $\mathcal{H}(\ln(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})))$. It is expressed as follows for all $(l,m) \in [\![1,d]\!]^2$ (the detailed steps are available in the Appendix B):

$$\mathcal{H}_{lm}(\ln(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})))(\boldsymbol{\theta}_i) = \begin{cases} \frac{\partial^2 \ln f_l(\theta_{il}|\boldsymbol{\Pi}_l)}{\partial \theta_{il}{}^2} - \left[\frac{f_l'(\theta_{il}|\boldsymbol{\Pi}_l)\varphi(\varrho_{il})^2 - f_l(\theta_{il}|\boldsymbol{\Pi}_l)^2\varphi'(\varrho_{il})}{\varphi(\varrho_{il})^3}\right]\left(u_{ll}\varrho_{il} + \sum_{\substack{m=1\\m\neq r}}^{d} u_{lm}\varrho_{im}\right) - u_{ll}\left(\frac{f_l(\theta_{il}|\boldsymbol{\Pi}_l)}{\varphi(\varrho_{il})}\right)^2 & \text{if } l = m \\ -u_{lm}\frac{f_l(\theta_{il}|\boldsymbol{\Pi}_l) f_m(\theta_{im}|\boldsymbol{\Pi}_m)}{\varphi(\varrho_{il})\varphi(\varrho_{im})} & \text{else} \ , \end{cases} \tag{3.24}$$

with $\varrho_{il} \coloneqq \Phi^{-1}(F_l(\theta_{il}|\boldsymbol{\Pi}_l)), l \in [\![1,d]\!]$, $\varphi(\cdot)$ the PDF of a Gaussian univariate distribution with 0 mean and unit variance and $\varphi'(\cdot)$ its derivative. This semi-analytical expression enables to avoid the use of finite-differences thus achieving a better accuracy. It works either for dependent or independent parameters (in the case of independent parameters, the Hessian matrix is diagonal and all other terms except the second derivative of the log-PDFs of the marginals in Eq. (3.24) vanish). In addition, as this matrix is diagonal, only upper (or lower) triangular terms can be estimated to gain computation time.

Before studying the calibration results, we should first check whether the individuals parameters are estimated properly to ensure an accurate estimation of the population parameters. To study this point, we repeat the process already implemented in the first section 3.4.1 for 1,000 full population parameters drawn within the bounds defined in Table 3.2, we carry out the minimization

of function $g_i(\cdot)$ defined in Eq. (3.20) given the data generated for a repetition (here the 9^{th}). With the provided individual parameters, we can estimate the averaged relative error from Eq. (3.12) available in Table 3.22. This test is achieved with two different scenarios: $\mathcal{S}_2$ in which all parameters are independent while the second, $\mathcal{S}_4$ includes statistical dependencies (this point has already been investigated in the previous sections for scenarios $\mathcal{S}_1$ and $\mathcal{S}_3$). Remember that for both of them, the marginal of the Young's modulus is a Gumbel distribution. The same means, standard-deviations and ω's are used in scenarios $\mathcal{S}_2$ and $\mathcal{S}_4$.

Table 3.22: For the 9^{th} repetition, averaged relative errors in % on the individual parameters $\mathbf{e}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_{50})$ in scenarios $\mathcal{D}_2$ and $\mathcal{D}_4$ for 1,000 proposals of full population parameters $\boldsymbol{\Psi}$ sampled between the bounds defined in 3.2.

	E_0	y_{0s}	y_{cs}	d_c
Averaged relative error on the individual parameters in scenario $\mathcal{S}_2$ [%]	8.93	88.9	20.9	12.9
Averaged relative error on the individual parameters in scenario $\mathcal{S}_4$ [%]	11.4	91.5	27.6	15.3

In Table 3.22, for scenario $\mathcal{S}_2$ (without dependency structure), the individual parameters appear to be estimated with the same accuracy as in Tables 3.3 and 3.8 for most proposal of full population parameters $\boldsymbol{\Psi}$. However, with statistical dependencies (scenario $\mathcal{S}_4$), it seems that they are estimated with increased error, especially the damage kinetics y_{cs}. This can be explained as follows: the correlations coerce the individual parameters to follow a given pattern, especially y_{cs} that is involved in two dependency relations. Consequently, the identification of E_0, y_{cs} and d_c will be harder to conduct for candidate distributions that are not consistent with the exact one, compared to the situation where no correlations are considered (scenario $\mathcal{S}_2$). Yet, the increase of the average relative errors remains limited and does not seem large enough to prevent a proper calibration of the population parameters. To terminate this validation, we can take a look at the convergence of the minimization of function $g_i(\cdot)$ given $\boldsymbol{\Psi} = \widehat{\boldsymbol{\Psi}}$ for the 9^{th} repetition, as depicted in Figure 3.31.

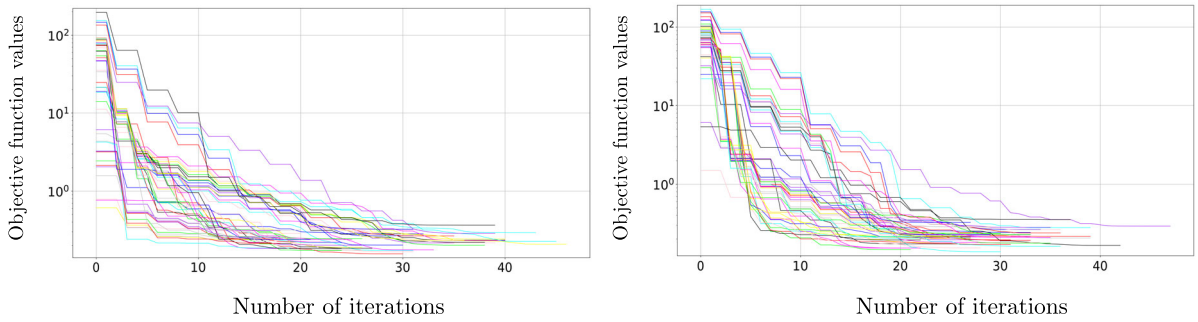


Figure 3.31: Objective function values $g_i(\cdot), i \in [\![1, 50]\!]$ for the 9^{th} repetition given $\boldsymbol{\Psi} = \widehat{\boldsymbol{\Psi}}$ (the MLE estimate). On the left, $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ for scenario $\mathcal{S}_2$ and on the right, $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ for scenario $\mathcal{S}_4$. Here, at a given iteration, only the best objective value out of the 10 restarts is plotted.

The minimization of $g_i(\cdot)$ does not seem to exhibit any particular issues. Thus, the estimation of the individual parameters is performed correctly and does not prevent likelihood maximization to find the MLE estimate as either depicts Figure 3.32 that displays, for all repetitions, the opposite log-likelihood values taken by the minimization algorithm.

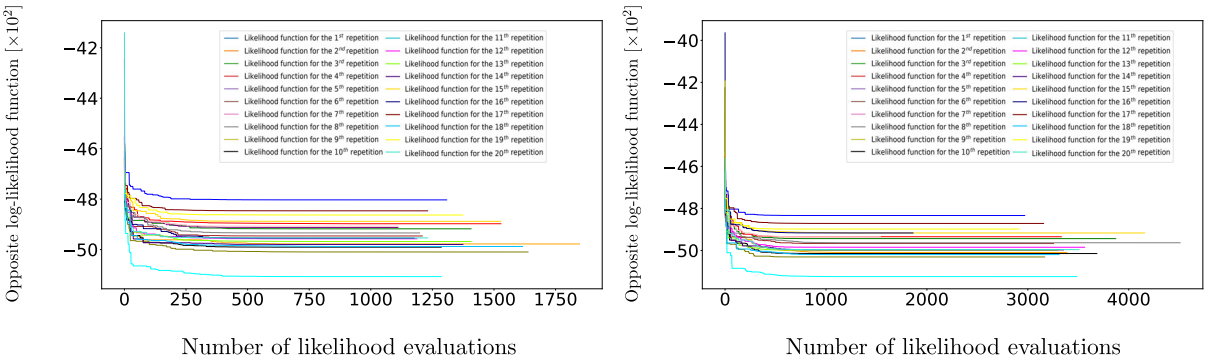


Figure 3.32: Best value of the opposite log-likelihood function values for scenario $\mathcal{S}_2$ on the left and on the right for scenario $\mathcal{S}_4$.

Briefly, we can see that the likelihood optimization for both scenarios $\mathcal{S}_2$ and $\mathcal{S}_4$ processes without major issues, allowing to proceed to the analysis of the results.

3.4.5.2 Examining the calibration for the different scenarios

The minimization is performed with the same settings (*i.e.*, starting point, maximum number of iterations, size of the population of the CMA-ES algorithm). Mixed-effects methods were applied with the data generated in the last section, that is to say generated from the ODM model with parameters following a Gaussian distribution with exact means and standard-deviations as provided in Table 3.1 and dependency structure characterized by two correlations: $\rho(d_c, y_{cs})$ equal to -0.4 and $\rho(d_c, y_{cs})$ to 0.8. To study the calibration results, we can first take a look to the marginals reported in Table 3.23.

Table 3.23: Comparison of the averaged error $\mathcal{E}_{\text{mean}}(\widehat{\mathbf{\Pi}})$ in % on the marginal parameters $\mathbf{\Pi}$ over 20 repetitions of the calibration process with samples of 50 specimens for 4 choices of model parameters distribution (scenarios $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ and $\mathcal{S}_4$). The coefficients of variation COV in % are indicated between brackets.

	E_0	y_{0s}	y_{cs}	d_c
Averaged errors on calibrated means with scenario $\mathcal{S}_1$ [%]	1.28 (81.0)	4.73 (62.4)	3.68 (76.4)	3.63 (72.2)
Averaged errors on calibrated means with scenario $\mathcal{S}_2$ [%]	1.17 (73.6)	4.70 (52.3)	3.50 (75.0)	3.49 (70.8)
Averaged errors on calibrated means with scenario $\mathcal{S}_3$ [%]	1.25 (67.4)	4.24 (74.4)	2.83 (63.8)	2.83 (84.4)
Averaged errors on calibrated means with scenario $\mathcal{S}_4$ [%]	1.26 (60.0)	4.27 (70.0)	2.91 (59.8)	2.80 (81.3)
Averaged errors on calibrated standard-deviations with scenario $\mathcal{S}_1$ [%]	10.6 (81.8)	34.1 (75.4)	13.6 (51.8)	10.6 (80.1)
Averaged errors on calibrated standard-deviations with scenario $\mathcal{S}_2$ [%]	16.2 (96.3)	28.8 (110)	13.9 (67.7)	10.8 (75.4)
Averaged errors on calibrated standard-deviations with scenario $\mathcal{S}_3$ [%]	12.0 (74.4)	24.8 (103)	12.7 (63.5)	8.33 (87.5)
Averaged errors on calibrated standard-deviations with scenario $\mathcal{S}_4$ [%]	13.8 (96.3)	20.2 (96.8)	12.2 (68.6)	8.50 (82.3)

In Table 3.23, the choice of different distributions does not seem to downgrade the estimation of the marginal of E_0. Indeed, for all scenarios, the average relative errors on its mean is similar, irrespectively of the choice of marginals or dependency structure. In fact, here, the

mean parameter that can be understood as a position parameter, is not influenced by the shape of variability. This remark suggests that, with the studied modelings, the fixed-effects are always properly estimated, even for an erroneous choice of model parameters distributions. Still, this conclusion could have changed if distribution with dramatically different shapes had been selected (for instance the arcsine distribution). The error on the standard-deviations also exhibit similar values, even if it is slightly higher when the Gumbel marginal is chosen. To explain this difference, we should remember that the standard-deviation characterizes the dispersion but also the shape of the distribution. Provided that the Gumbel distribution is a non-symmetric distribution unlike its Gaussian counterpart, the increased error for the standard-deviations comes from the fact that the shape of the assumed distribution (Gumbel) does not fit the shape of the distribution used to generate the exact individuals (which is Gaussian). Then, to best fit the empirical distribution of the individual parameters, one solution is to bias the estimation of the standard-deviation that results in an increased error. In addition, the estimation of the other standard-deviations may be slightly downgraded by at most 15% from scenarios $\mathcal{S}_3$ and $\mathcal{S}_4$ to $\mathcal{S}_1$ and $\mathcal{S}_2$. To understand this point, we should remember that, as recalled earlier, the standard-deviation parameter controls the dispersion of the marginal and somehow implement constraints on the dispersion of the realizations, similarly to the statistical dependencies. Then, if we assume independence between the model parameters, the only available degrees of freedom to set up constraints are the standard-deviation parameters that combine both informations together and are thus altered from the true modeling. In fact, choosing an erroneous distribution for E_0 (or dependency structure for scenarios $\mathcal{S}_1$ to $\mathcal{S}_3$) does not appear to have a large impact on the calibration of the other marginals. To summarize the observations, it seems that considering an inexact assumption on the marginal distributions and the dependency modeling does not influence the estimation of the location parameters, but downgrades the calibration of the shape and dispersion parameters. Still, we should insist on the fact that with our choices of modeling, this loss of accuracy does not completely flaw the estimation of the population parameters that remain estimated with reasonable precision (at most 15% for the dispersion parameters that require, whatever assumption, more individuals to achieve the same accuracy as for the location parameters). Finally, though the marginal of y_{0s} seems more impacted by the modeling assumptions, the low sensitivity of the model output with respect to this parameter prevent to draw clear conclusions about it. We can only say that it reaches acceptable level of errors given the difficulty to estimate accurately its marginal. For the 9$^{\text{th}}$ repetition, the calibrated marginals for 4 different scenarios alongside the corresponding individual parameters are displayed in Figure 3.33.

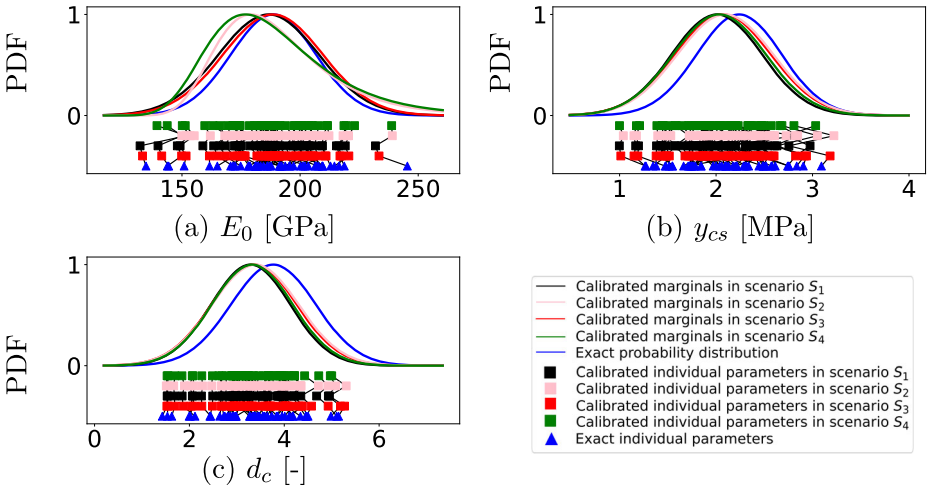


Figure 3.33: Exact and calibrated marginals for all scenarios along the E_0-axis (a), the y_{cs}-axis (b) and the d_c-axis (c) for the 9$^{\text{th}}$ repetition. Individual parameters are also plotted.

In Figure 3.33, the remarks from above are confirmed. For all calibrated marginals, it can be noticed that the support of the calibrated distribution for which the associated PDF is non-zero matches with the intervals containing the individual parameters. Note the presence of 4 saturations lower than 2. Those specimens, outside the 95% confidence intervals, can be considered as outliers and their presence partly explains the increase of the standard-deviations. This point is developed further below. Furthermore, it seems that the estimation of the individual parameters provide the same estimation for the different configurations of $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$. This topic will be examined further below. It is now possible to investigate the calibration of the statistical dependencies (here in scenarios $\mathcal{S}_3$ and $\mathcal{S}_4$). The values of interest are communicated in Table 3.24.

Table 3.24: For 20 repetitions with 50 specimens each, averaged calibrated correlations $\widehat{\boldsymbol{\Pi}}_{\text{mean}}$ and associated averaged relative error $\mathcal{E}_{\text{mean}}(\widehat{\boldsymbol{\Pi}})$ in % for scenarios $\mathcal{S}_3$ and $\mathcal{S}_4$. The corresponding coefficients of variation COV in % are indicated between brackets.

	$\rho(E_0, y_{cs})$	$\rho(d_c, y_{cs})$
Exact values	-0.4	0.8
Average calibrated correlations for scenario $\mathcal{S}_3$	-0.370 (41.0)	0.798 (8.00)
Averaged errors on the correlation parameters for scenario $\mathcal{S}_3$ [%]	30.7 (76.7)	6.65 (66.4)
Average calibrated correlations for scenario $\mathcal{S}_4$	-0.394 (40.8)	0.791 (8.97)
Averaged errors on the correlation parameters for scenario $\mathcal{S}_4$ [%]	30.8 (84.2)	6.80 (87.4)

In Table 3.24, we can notice that both correlations are estimated with equivalent degrees of precision in both scenarios, even $\rho(d_c, y_{cs})$ that involves E_0 for which we try different marginals. To be more precise, on average, $\rho(E_0, y_{cs})$ is better calibrated in scenario $\mathcal{S}_4$ than in scenario $\mathcal{S}_3$, contrary to the other correlation. In fact, the small differences between the different indicators come more likely from numerical errors rather than structural reasons, suggesting that the estimation of the correlations is not flawed if the marginals are misspecified. Still, notice that both the averaged errors and their associated COVs with scenario $\mathcal{S}_4$ are higher than those with scenario $\mathcal{S}_3$, indicating a greater variability for the estimation of correlation coefficients which advocates for poorer estimations for few repetitions. In such cases, it is difficult to decide whether this come from a numerical issues (for the Hessian calculation for instance) or a consequence of an erroneous choice for $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$. Furthermore, we even find the same pattern for the correlations, that is to say $\rho(d_c, y_{cs})$ is better estimated than $\rho(E_0, y_{cs})$ because the latter is less selective (so the likelihood function less sensitive) than the first. To sum it up, the choice of different marginals does not seem to flaw the estimation of the dependency structure in this test case. Figure 3.34 illustrates the calibration of the statistical dependencies for both scenarios and confirms the remarks made here.

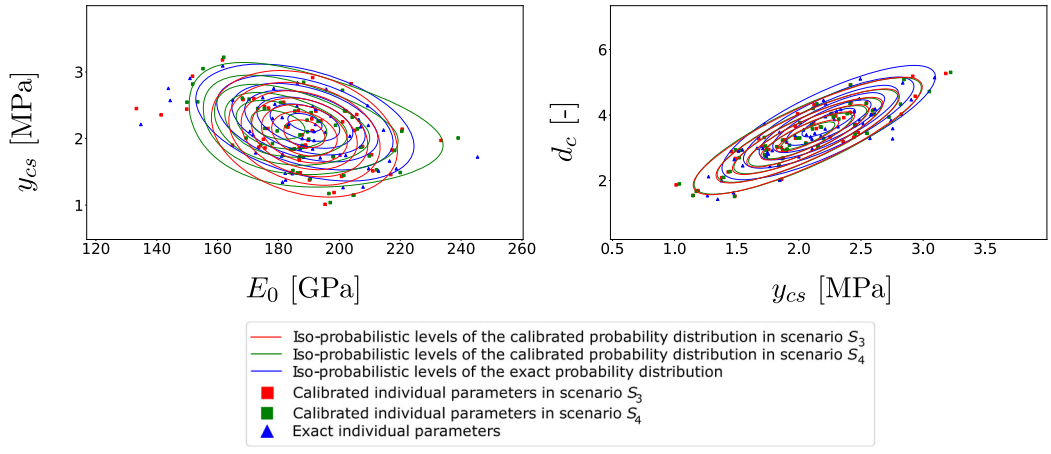


Figure 3.34: Exact and calibrated dependency structure for scenarios $\mathcal{S}_3$ and $\mathcal{S}_4$ for the 9$^{\text{th}}$ repetition. Individual parameters are also plotted.

To summarize the results, we can study joint criteria that respectively refer to KL divergence and the Bayesian Information Criterion (BIC). Both of them intend to test the quality of the results in different ways. The first studies how close the calibrated distribution is to the exact distribution while the latter encompasses the adequation of the $\boldsymbol{\theta}_i$s to $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$, the correspondence between the experimental data $\mathbf{y}_i$ to the model for the $\widehat{\boldsymbol{\theta}}_i$s and the number of population parameters. Both of them are depicted in Figure 3.35.

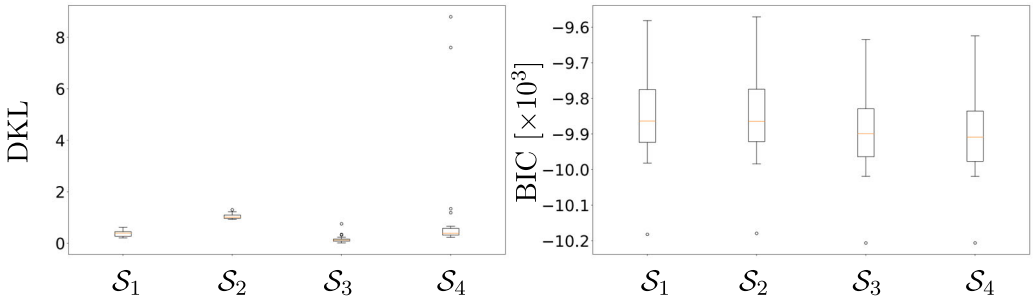


Figure 3.35: For all repetitions, KL divergence between the exact and calibrated distribution on the left and BICs on the right (the lower the better).

On the left of Figure 3.35, we can see that the estimated distributions for scenario $\mathcal{S}_2$ refer somehow to the poorest calibrations given their KL divergences values. This could be expected because the model parameters distribution in this scenario is based on the most erroneous assumptions: one of the marginals is false and it does not include a correlation structure. For intermediate situations (scenarios $\mathcal{S}_1$ and $\mathcal{S}_4$), we can notice a significant improvement in the estimation of the model parameters, without being able to state which is the closest (it depends if we focus on the medians, on the maximum gap, *etc.*). Finally, as it could be expected, it is with the exact assumptions that the estimated distribution approximates the best the exact distribution. However, in practice, the exact shape of the model parameters distribution is unknown (and computing KL divergence is not possible) but the BIC criterion can be estimated, as on the right plot of Figure 3.35. The BIC criterion, which depends on the log-likelihood and is penalized by the number of population parameters including those of the Gaussian copula (so 9 in scenarios $\mathcal{S}_1$ and $\mathcal{S}_2$ and 11 in scenarios $\mathcal{S}_3$ and $\mathcal{S}_4$), seems to be less conclusive. It seems that to improve the calibration, the decisive step is the inclusion of a dependency structure for both choices of marginals, that are themselves comparable. Changing the marginal of E_0 does not change the likelihood which is consistent with previous observations (in particular from Table 3.23). Indeed, here, the improvement of the likelihood function when correlations are considered outranks the increase of the penalization of the BIC formulation (because of the extra parameters), comforting this diagnosis. However, the difference are relatively small: from scenarios $\mathcal{S}_1$ and $\mathcal{S}_2$ from scenarios $\mathcal{S}_3$ and $\mathcal{S}_4$, the improvement is smaller than 1%. To conclude, though the calibration that includes statistical dependencies provides better results in terms of BICs, the differences remain low between the different scenarios and are relevant provided the available data. One reason that can explain the low impact of the change of marginal is that, even if the Gumbel distribution is not symmetric contrary the Gaussian one, their associated PDFs have the same smoothness properties. To detect a significant change, the Laplace distribution as in [Demidenko, 2013] may have been more suited, though it is rarely used in mechanical engineering. In fact, the indicators presented do not seem able to discriminate between the different scenarios. To choose between the different scenarios, in other words the choice of $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$, model testing could be carried out using the appropriate statistical test developed by [Drikvandi, 2017]. Nevertheless, it is computationally prohibitive as it implies both Monte-Carlo sampling allied with bootstrap. Furthermore, with few individuals (from a statistical point of view), clear conclusions may be hard to draw and is therefore not implemented.

Figure 3.35 already gave us clues on the calibration of the individual parameters. Indeed, the BIC is largely based on the likelihood, that combines a global approach (population fit) and a local one (individual fit). Provided that the ranges of variation are similar for all scenarios, it seems that the error on the individual parameters is comparable for all scenarios. This is confirmed with a more appropriate indicator, the average relative errors on the $\boldsymbol{\theta}_i$s reported in Table 3.25.

Table 3.25: Comparison of the averaged errors on the individual parameters $\mathbf{e}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_{50})$ in % over 20 repetitions of the calibration process with samples of 50 specimens each for 4 choices of model of model parameters distribution (scenarios $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ and $\mathcal{S}_4$). The coefficients of variation COV in % are indicated between brackets.

	E_0	y_{0s}	y_{cs}	d_c
Averaged errors on the estimated individual parameters with scenario $\mathcal{S}_1$ [%]	2.98 (14.1)	30.2 (47.4)	8.42 (8.48)	4.47 (9.16)
Averaged errors on the estimated individual parameters with scenario $\mathcal{S}_2$ [%]	3.06 (13.9)	29.0 (43.2)	8.40 (7.71)	4.40 (8.76)
Averaged errors on the estimated individual parameters with scenario $\mathcal{S}_3$ [%]	2.83 (17.2)	23.4 (43.1)	6.76 (16.4)	3.82 (17.2)
Averaged errors on the estimated individual parameters with scenario $\mathcal{S}_4$ [%]	2.80 (17.9)	22.4 (38.6)	7.01 (14.7)	3.81 (16.4)

In Table 3.25, it appears that for all scenarios, the individual parameters seem to be properly estimated (except y_{0s} for reasons already exposed earlier). The level of accuracy is equivalent for all scenarios, which is reassuring because, if the assumption on the $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ is somehow arbitrary, the $\boldsymbol{\theta}_i$s originate from the data that does not change from a scenario to the other. Still, it can be noticed that the COVs of the errors are higher when statistical dependencies are considered, hinting that the level of error on the $\boldsymbol{\theta}_i$s can reach significant values for some of the repetitions. To some extent, it can be explained by the fact that the inclusion of a dependency structure coerces the individual parameters to comply with the pattern of $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ that may sometimes conflicts with the data itself. This agreement can be illustrated by representing for all scenarios and one repetition (the second), the exact individual parameters against the exact ones as in Figure 3.36.

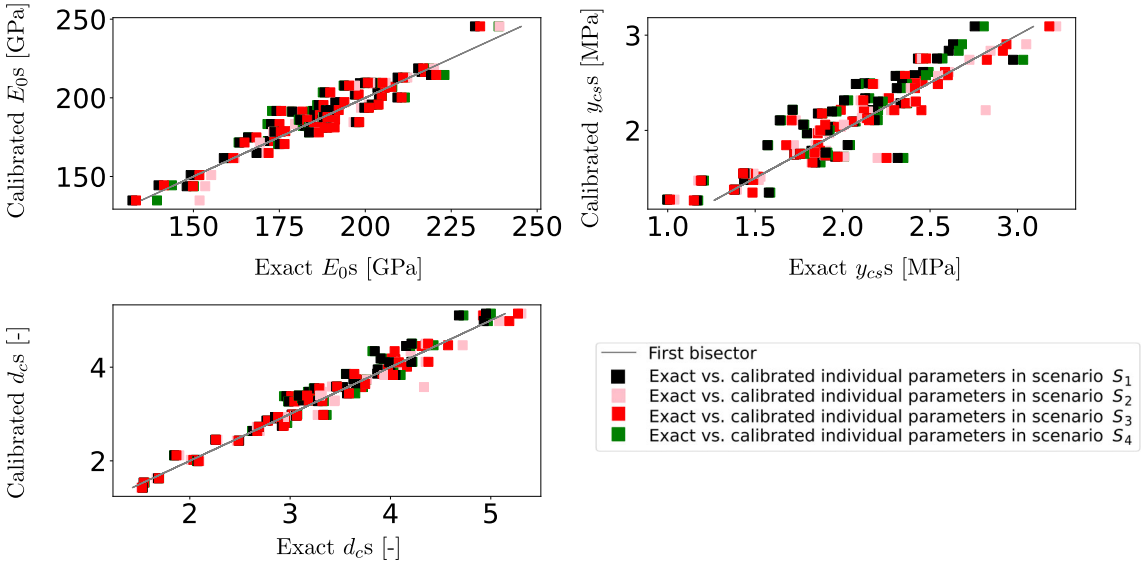


Figure 3.36: For the 9$^{\text{th}}$ repetition, exact versus calibrated individual parameters for the E_0 (top left), y_{0s} (top right), y_{cs} (bottom left) and d_c (bottom right).

Figure 3.36 depicts the good overall agreement the exact and calibrated individual parameters, even if for some specimens, the estimations can sometimes be inconsistent with the exact values. This is consistent with the fact that the individual kinetics are estimated with more errors than the other parameters. It is represented by the fact that in Figure 3.36, the data points are more dispersed around the first bisector than the individual saturations that are more concentrated around it. The accurate estimation of the individual parameters provides an accurate estimation of the stresses, as reported in Table 3.26. Note the weak COVs of the errors (below 10%), which shows that all individuals are properly estimated.

Table 3.26: Averaged distance $D(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_{50})$ in model space between the exact and calibrated models for the 50 individual parameters averaged over 20 repetitions of the calibration process for all scenarios. The coefficients of variation COV in % are indicated between brackets.

	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_4$
Averaged error in model space [MPa]	1.91 (6.56)	1.90 (6.72)	1.84 (6.80)	1.85 (6.63)

3.4.5.3 Comparing the estimations in model space

So far, we have studied the estimated variability by checking the error on the different population parameters. Nevertheless, in mechanical design, quantities of interest are often in model space. Thus, it is of prime importance to check, though the population parameters are estimated with different assumptions, whether the estimations of variability in the design space are comparable. To evaluate this point, we propagate the different distributions through the ODM model. For scenario $\mathcal{S}_4$, the generation requires the use of an iso-probabilistic transformation [Rosenblatt, 1952]. To generate the samples, we first generate samples from $\mathcal{N}(0_{\mathbb{R}^4}, \widehat{\mathbf{R}})$, with $\widehat{\mathbf{R}}$ the estimated correlation matrix. Then, for each of the coordinates $l \in [\![1, d]\!]$, we apply transform $F_l^{-1} \circ \Phi(\cdot)$. The first part of the transform ($\Phi(\cdot)$) maps the Gaussian distributed numbers with a 0 mean and a unit variance into uniform numbers and the second ($F_l^{-1}(\cdot)$) maps the uniform numbers into the standard space. In fact, for all parameters but the first, $F_l^{-1} \circ \Phi$ reduces to the linear function $x \mapsto \mu_{\theta_l} + \mathrm{sd}(\theta_l)x$, with μ_{θ_l} and $\mathrm{sd}(\theta_l)$ the mean and standard-deviations of the l^{th} model parameter. From the model outputs computed from the 1,000 samples generated from $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$, we can, for each observation time $(t_{ij})_{(i,j)\in[\![1,n]\!]\times[\![1,N_i]\!]}$, estimate a 95% confidence interval as the difference between the 2.5% and 97.5% quantiles, as depicted in Figure 3.37.

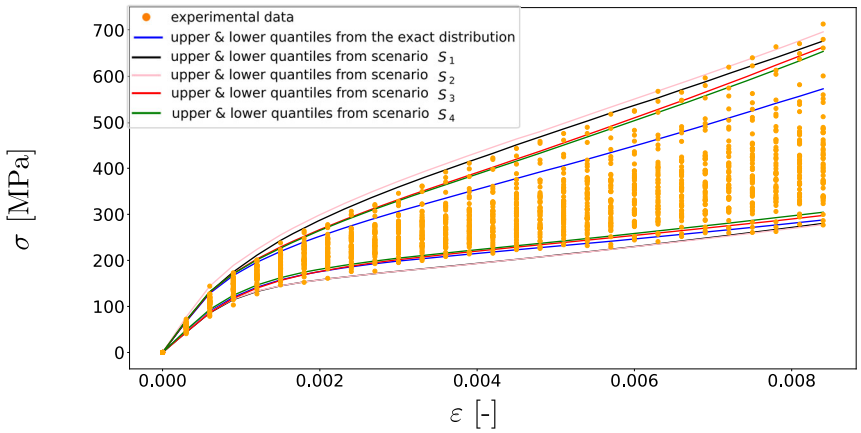


Figure 3.37: For the 9^{th} repetition, uncertainty propagation with 1,000 samples of the exact and calibrated distributions for the 4 different scenarios, along the experimental data.

In Figure 3.37, we can see that for all scenarios, the estimation of the material variability in model space is consistent with the experimental data. It can be noticed that the bundles for scenarios that do not consider dependency modeling (scenarios $\mathcal{S}_1$ and $\mathcal{S}_2$) are wider compared to the other scenarios. This is a direct consequence of the presence of correlations between model

parameters that implement constraints between them, and thus reduce the observed variability. The higher quantiles estimated for the distributions that do not include correlations seems more relevant than those estimated with correlations included as they encompass more experiments. Still, it should be noted that the individuals above the higher quantile of the exact distribution refer in fact to the specimens highlighted in Figure 3.33 that can be considered as outliers as the associated saturations are outside their corresponding 95% confidence interval. However, in the calibration process, this information is not known and these specimens have the same impact on the likelihood function as the other specimens. The calibration therefore accounts for them as it does for the other experiments and enlarges the standard-deviation of the damage saturation, resulting in larger bundles for the calibrated distributions. The lower quantiles do not follow the same pattern and seem to be the same for the different scenarios. This could be explained by the presence of a lower bound for the model outputs that limits the space of the physical outputs common to all scenarios. On the contrary, no upper bound seems to appear, which can explain why the difference between the quantile curves expresses more clearly. Finally, we can notice that in this case, the choice of the marginals does not strongly influence the propagated variability, at least not as much as the correlations.

3.4.6 Brief recap of the results

In this section, we studied the calibration of the ODM model with mixed-effects using synthetic data. From a first implementation with few individuals and the CMA-ES algorithm for the auxiliary optimization, we showed that the population and individual parameters could be identified with a good accuracy. Then, we discussed the choice of the algorithm for the auxiliary optimization that was replaced by a multi-start SLSQP to gain computational time with the same precision of the results. We investigated the influence of the number of individuals to show that adding individuals both reduces error and uncertainty on the population parameters, at the expense of an increase in both experimental and computational costs, even if the latter do not mobilize the same amount of financial resources (the cost of the experimental campaigns is usually higher than the cost of the computational ones). The calibration of a dependency model was studied, with a focus on the interest of adding such assumptions with few individuals, showing in this case that the estimation of the low-level statistical dependencies can be flawed with much error ,making it more interesting to ignore small correlations than to evaluate them. Finally, we successfully explored calibration with an other assumption than the usual Gaussian hypothesis (replaced with a Gumbel distribution). This study showed that, because of the comparable shapes of the tested marginals for E_0, only slight deteriorations of the population parameters could be detected. However, at the level of the complete joint distribution, the studied indicators estimated either in model or parameter space in this test-case did not really enable to discriminate the choice of the marginals. This thorough investigations show that mixed-effects can find in practice the population and individual parameters, in the context of behavior law calibration with experiments from which specimen variability arises, provided that the number of individuals is sufficient. It is possible to go through calibration of the ODM model with real data.

3.5 Calibrating from real data

The previous sections have allowed us to investigate mixed-effects calibration through different points of view. In particular, we saw that the approach could estimate model parameters distribution outside the usual Gaussian framework and that statistical dependencies between the model parameters could be added. We seek now to process with mixed-effects the data presented in Figure 3.1. This application is of prime interest because the real data presents challenges that are not present with synthetic data. For instance, the data collected on the

testing machine in the form of load-displacement curves has to be processed to give stress-strain curves, which may introduce bias and increase the levels of the measurements errors. On the contrary, the synthetical data is directly generated from the model and does not require any specific treatment. Unexpected events can also alter the data such as sudden failure because of poor experimental set up. Here, Section 3.5.1 will briefly remind the data that will be used in the calibration process before describing the calibration settings (with the optimization) in Section 3.5.2 to finally report the results from mixed-effects calibration for the ODM model in Sections 3.5.3, 3.5.4 and 3.5.5.

3.5.1 Data presentation

The mixed-effects method is now applied to calibrate the ODM model with 13 real tensile tests performed on the CERASEP A400 woven composite material. The tests measures are plotted in Figure 3.38. The curves are less disparate than the synthetic data of Figure 3.2. The transition between the linear and nonlinear regimes is subject to a significant variability constituting an interesting test case for a population-based approach such as the mixed-effects models, even if the curves tend to overlap when damage increases, for strains superior to 0.4 %. Notice also that the strain-stress curves have different numbers of points (from 20 for 8 of the experiments to 700 points). Consequently, to make sure that all individuals have the same weight in the likelihood function, strain-stress curves with more than 20 observation points are subsampled periodically and always include the last point. Yet, this sub-sampling reduces the amount of available information, thus increasing the level of measurements errors which still remain negligible with respect to material variability. All curves considered in the following have 20 observation points.

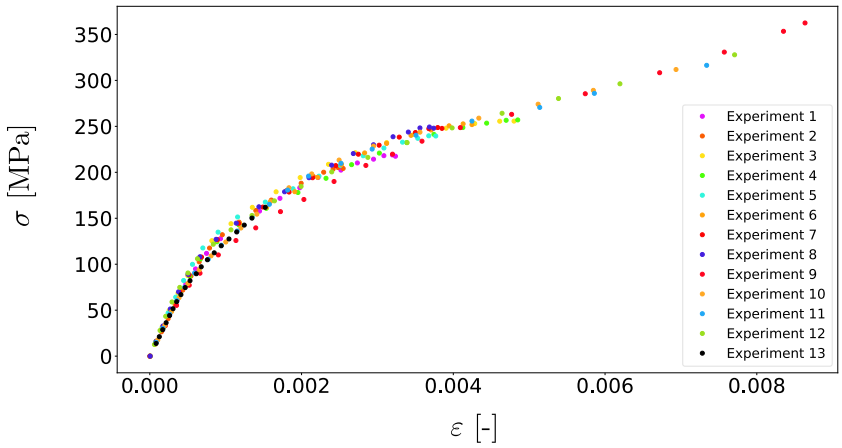


Figure 3.38: 13 experimental monotonic tensile strain-stress curves of a Ceramic Matrix Composite. Compared to Figure 3.1, the strain-stress curves are subsampled.

In Figure 3.38, the failure of the 13^{th} experiment occurs at 0.2% strain, that is to say much sooner than any of the other 12 repetitions (for which it happens after 0.4% strain). This experiment that breaks much earlier than the others suggests that a possible issue took place during the experiment. Moreover, we can notice that nonlinear behavior does not fully express, prompting to carefully study the behavior of the identification of the individual parameters for all experiments and the last especially. This is included in the next which precises several of the settings for the minimization of the log-likelihood of mixed-effects.

3.5.2 Calibration settings

To start this study, it is necessary to define the assumptions on the statistical model (*i.e.*, on the marginals, copulas, residual model) to determine the population parameters to estimate, and set up the optimization problem of the likelihood. The hypotheses that will be considered

here will consist into those studied in the previous sections, that is to say Normal and Gumbel marginals with independent and Gaussian copulas. So far, we have supposed that the variance of the measurements errors (a Gaussian white noise) was the same for all experiments and observations. This assumption was acceptable because the same residual model was used to generate the synthetic data. A more realistic scenario is to suppose different residual models for each of the available individuals. This is consistent with the observations in Chapter 1 where the experiments exhibited similar but different levels of noise. Still, for the sake of completeness, one calibration will be carried out with a single residual model for all experiments, but only with an independent copula and Gaussian marginals. In the dependent case, all dependencies except those involving y_{0s} will be considered because this parameter is complicated to estimate. In total, this makes 5 calibration scenarios, all of them described with the different hypotheses and number of parameters in Table 3.27.

Table 3.27: Different assumptions for the calibration of the ODM model with mixed-effects alongside their associated colors.

Scenario	Marginals	Copula	Residual model	Number of parameters to estimate
$\mathcal{D}_1$	Gaussian	Independent	Single model error for all individuals	9
$\mathcal{D}_2$	Gaussian	Independent	One model error for each individual	20
$\mathcal{D}_3$	All Gaussian except a Gumbel marginal for E_0	Independent	One model error for each individual	20
$\mathcal{D}_4$	Gaussian	Gaussian except for y_{0s}	One model error for each individual	23
$\mathcal{D}_5$	All Gaussian except a Gumbel marginal for E_0	Gaussian except for y_{0s}	One model error for each individual	23

To ensure the consistency of the results, calibration process is repeated several times. However, because of the computational costs, only 4 repetitions are considered. To preserve the consistency between the numerical indicators studied with synthetical and real data, the initialization points in this section refer to those used in Section 3.5.4. The same optimization settings (number of iterations, population size, *etc.*) are kept, except we have to define new bounds, described in Table 3.28. For all scenarios, the bounds are the same, but can be slightly changed if it proves to be necessary (for instance if one of the population parameters hit on the bounds). In such a case, the calibrations are re-started with the updated bounds. For the scenarios that imply different residual models for each individual, the bounds on standard-deviation of the noise remain the same (hence the index (i)). The bounds on the population parameters are determined from the empirical means and standard-deviations evaluated after the independent calibration of each of the available specimens. In general, the bounds for the mean parameters determined this way are appropriate but those on the standard-deviations are often too high, because the independent calibrations do not account for the intrinsic variability of the material properties. The bounds on $\omega_{(i)}$ are determined by trial-error.

Table 3.28: Search space for the population parameters.

μ_{E_0} [GPa]	$\mu_{y_{0s}}$ [kPa]	$\mu_{y_{cs}}$ [MPa]	μ_{d_c} [-]
$[120, 235]$	$[3.00, 50.0]$	$[2.10, 4.50]$	$[1.70, 6.50]$

$\mathrm{sd}(E_0)$ [GPa]	$\mathrm{sd}(y_{0s})$ [kPa]	$\mathrm{sd}(y_{cs})$ [MPa]	$\mathrm{sd}(d_c)$ [-]	$\omega_{(i)}$ [MPa]
$[1.40, 38.0]$	$[1.00, 12.0]$	$[10^{-2}, 3.50]$	$[5.00 \times 10^{-2}, 2.25]$	$[0.707, 7.07]$

Before proceeding to the results, it must be checked whether the individual parameters are properly identified. As there is no exact value for the individual parameters, the only possibility

left is to compute an Root Mean Square Error (RMSE) between the experimental data and calibrated model. To assess this point, for all scenarios, the accuracy of the calibration is examined. For 1,000 $\boldsymbol{\Psi}$ uniformly drawn within the bounds defined in Table 3.28, the individual parameters for 2 experiments (the twelfth and the last one) are estimated. These repetitions are chosen because failure occurs much earlier for the thirteenth sample than for the others which can be embodied by any other experiment such as the twelfth. and this for all scenarios as depicted in Figure 3.38.

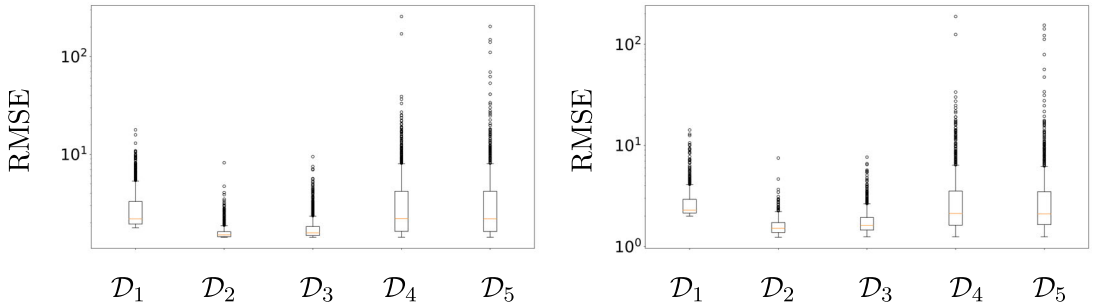


Figure 3.39: For the twelfth (on the left) and thirteenth experiments (on the right), RMSEs between the experimental and calibrated curves for the individual parameters corresponding to 1,000 $\boldsymbol{\Psi}$ uniformly drawn within the bounds defined in Table 3.28.

Figure 3.39 shows that, for all scenarios, the specimens are estimated with reasonable accuracy. It shows that the RMSEs remain lower than 10 MPa when the stress value goes well over 100 MPa (see Figure 3.1). Yet, some of the calibrations seem to provide poor estimations of the experimental stress-strain curves. However, these calibrations refer to population parameters that are not consistent with the experimental data. This phenomenon is more pronounced when correlations are included, which is consistent with earlier observations from Table 3.22. Another verification that can be carried out consists in evaluating the variability of the estimation of the individual parameters to make sure that the information used to determine the individual parameters mainly comes from the data and not the regularization term that is imposed through the assumption on $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$. To test this point, from the 1,000 $\boldsymbol{\Psi}$s sampled previously, the COV of the estimation of the individual parameters for the same repetitions is computed and available in Table 3.29.

Table 3.29: For 1,000 $\boldsymbol{\Psi}$ uniformly drawn within the bounds of Table 3.28, COVs of the calibrated individual parameters for the 12^{th} experiment $\widehat{\boldsymbol{\theta}}_{12}$ and 13^{th} experiments $\widehat{\boldsymbol{\theta}}_{13}$ for all scenarios.

	COV(E_0) [%]	COV(y_{0s}) [%]	COV(y_{cs}) [%]	COV(d_c) [%]
Experiment 12 in scenario $\mathcal{D}_1$	10.6	70.3	25.9	17.6
Experiment 12 in scenario $\mathcal{D}_2$	8.77	81.1	25.4	13.0
Experiment 12 in scenario $\mathcal{D}_3$	7.23	82.6	22.5	12.1
Experiment 12 in scenario $\mathcal{D}_4$	17.4	75.6	50.8	28.4
Experiment 12 in scenario $\mathcal{D}_5$	16.2	79.9	48.7	28.1
	COV(E_0) [%]	COV(y_{0s}) [%]	COV(y_{cs}) [%]	COV(d_c) [%]
Experiment 13 in scenario $\mathcal{D}_1$	5.33	80.0	42.0	24.6
Experiment 13 in scenario $\mathcal{D}_2$	3.32	52.9	84.8	47.9
Experiment 13 in scenario $\mathcal{D}_3$	5.58	58.9	85.9	47.2
Experiment 13 in scenario $\mathcal{D}_4$	7.79	77.5	87.6	52.8
Experiment 13 in scenario $\mathcal{D}_5$	9.70	79.2	87.0	52.3

Let's first focus on $\boldsymbol{\theta}_{12}$ which characterizes the twelfth experiment. Table 3.29 shows that the for many different candidates of population parameters, the estimated individual parameter, $\widehat{\boldsymbol{\theta}}_{12}$, remains more or less the same, due to its limited range of variation. As the population parameters are not held fixed, this variation suggests that this value is mainly driven by the experimental data but remains impacted by the regularization term monitored by the population parameters. It works for all scenarios, though when adding correlations between the parameters, the estimated values exhibit higher variations. It indicates that the choice of $\boldsymbol{\Pi}$ seems to more decisive, which can be understood as adding a dependency structure to $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ constrains the $\boldsymbol{\theta}_i$s to follow a pattern that may not match with the data. On the contrary, the estimated individual parameters of the 13$^{\text{th}}$ specimen $\widehat{\boldsymbol{\theta}}_{13}$ demonstrate greater variability compared to the identification $\widehat{\boldsymbol{\theta}}_{12}$ for the different proposal of $\boldsymbol{\Psi}$s. It therefore follows that $\widehat{\boldsymbol{\theta}}_{13}$ is more dependent on $\boldsymbol{\Psi}$, so on the choice of $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ than $\widehat{\boldsymbol{\theta}}_{12}$. Indeed, with this individual and all scenarios, two of the parameters, namely y_{cs} and d_c turn out to be heavily dependent on the choice of $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$, as if within function $g_i(\cdot)$, the regularization term $(-\ln(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})))$ had more impact on the determination of these parameters than the fitting term $(-\ln(f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i)))$ that relies on the data. To confirm this diagnosis, we can plot, as proposed in Figure 3.40 for the two previous experiments the relative contribution $\varpi_i(\cdot)$ of the regularization term to the function $g_i(\cdot)$ that is expressed as $\varpi_i(\widehat{\boldsymbol{\theta}}_i, \mathbf{y}_i, \boldsymbol{\Psi}) := \frac{|\ln(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\widehat{\boldsymbol{\theta}}_i|\boldsymbol{\Pi}))|}{|\ln(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\widehat{\boldsymbol{\theta}}_i|\boldsymbol{\Pi}))|+|\ln(f(\mathbf{y}_i|\widehat{\boldsymbol{\theta}}_i,\boldsymbol{\Pi},\omega_i))|}$.

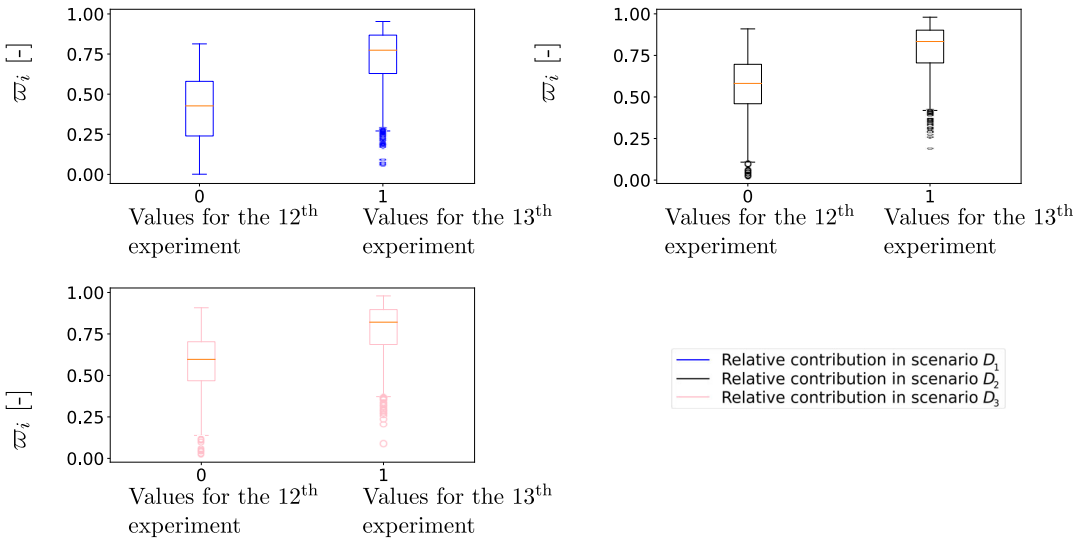


Figure 3.40: Relative contributions for 1,000 $\boldsymbol{\Psi}$ uniformly drawn within the bounds defined in Table 3.28 for the 12$^{\text{th}}$ and the 13$^{\text{th}}$ experiment in scenarios $\mathcal{D}_1$ (top left), scenario $\mathcal{D}_2$ (right) and scenario $\mathcal{D}_3$ (bottom left).

Figure 3.40 corroborates the fact that, on average, the weight of the regularization term within function $g_i(\cdot)$ is greater than the weight of the fitting term for the thirteenth experiment. From a physical point of view, it comes from the fact that both y_{cs} and d_c are involved in the modeling of damage that does not express much because of the early failure of this specimen, which explains their higher variability compared to the other specimens. On the contrary, as the 13$^{\text{th}}$ experiment mainly consists in an elastic linear behavior (approximately half of the points), the identification of E_0 exhibits smaller variability compared to the other experiments as it explains most of the experimental observations. This limits the interest of including this specimen within the database as this specimen does not add extra-knowledge with respect to the other experiments. It may even make harder the identification of the population parameters related to y_{cs} and d_c. Thus, in a first time, we choose not to include this experiment among the 12 others in the calibration process whose results are presented in the next section. The same analyses (on the population parameters, individual parameters, *etc.*) are carried out.

3.5.3 Calibration results with mixed-effects

We now calibrate the CERASEP A400 material with mixed-effects considering both dependent and independent model parameters. Of course, contrary to synthetic data, it is not possible to confront the population and the individual parameters estimates to their exact counterpart that are, by definition, unknown. In Table 3.30, the calibrated population parameters for the marginals are reported.

Table 3.30: Averaged calibrated marginal parameters $\widehat{\mathbf{\Pi}}_{\text{mean}}$ over 4 repetitions of the calibration process with 12 specimens each. The coefficients of variation COV in % are indicated between brackets.

Averaged calibrated means	E_0 [GPa]	y_{0s} [kPa]	y_{cs} [MPa]	d_c [-]
Scenario $\mathcal{D}_1$	203 (0.140)	0.158 (32.5)	4.19 (0.190)	5.23 (0.180)
Scenario $\mathcal{D}_2$	180 (0.204)	5.81 (6.08)	3.03 (1.06)	4.23 (0.622)
Scenario $\mathcal{D}_3$	180 (0.220)	6.11 (2.55)	2.99 (0.0596)	4.21 (0.0402)
Scenario $\mathcal{D}_4$	181 (0.0360)	5.39 (0.421)	3.14 (0.0677)	4.31 (0.0671)
Scenario $\mathcal{D}_5$	180 (0.179)	5.59 (3.37)	3.09 (1.26)	4.26 (0.653)
Averaged calibrated standard-deviations	E_0 [GPa]	y_{0s} [kPa]	y_{cs} [MPa]	d_c [-]
Scenario $\mathcal{D}_1$	17.6 (0.201)	0.409 (31.2)	1.34 (0.607)	0.807 (0.430)
Scenario $\mathcal{D}_2$	13.2 (0.138)	3.61 (3.67)	1.03 (0.830)	0.591 (0.663)
Scenario $\mathcal{D}_3$	15.3 (0.158)	3.78 (0.320)	1.00 (0.514)	0.570 (0.311)
Scenario $\mathcal{D}_4$	12.9 (0.397)	2.97 (0.108)	1.22 (0.656)	0.764 (0.727)
Scenario $\mathcal{D}_5$	15.0 (0.775)	3.11 (4.28)	1.19 (1.67)	0.722 (1.91)

Table 3.31: Averaged calibrated correlations $\widehat{\mathbf{\Pi}}_{\text{mean}}$ over 4 repetitions of the calibration process with 12 specimens each. The coefficients of variation COV in % are indicated between brackets. N/A stands for non-attributed.

Averaged calibrated correlations	$\rho(E_0, y_{cs})$	$\rho(E_0, d_c)$	$\rho(y_{cs}, d_c)$
Scenario $\mathcal{D}_4$	-0.392 (1.00)	-0.0249 (N/A)	0.892 (0.147)
Scenario $\mathcal{D}_5$	-0.429 (1.65)	-0.0304 (N/A)	0.880 (0.321)

A first remark that can be made on the estimation of the population parameters is to notice that the choice of the residual model has a major impact on the calibration results. Indeed, for all parameters and in particular the damage threshold, a difference up to 40% can be observed between the estimations that consider different residual models for each individual or a single one for all specimens. It is worth noticing that the estimation of the standard-deviations of E_0 depends on the choice of the marginal Gaussian or Gumbel marginal (the Gaussian distribution is symmetric contrary to the Gumbel one). Furthermore, for both y_{cs} and d_c, the standard-deviations estimated without dependency structure are always larger than when statistical dependencies are considered, illustrating the impact of the correlation parameters. Indeed, the correlation between those two is calibrated about 0.9 (in Table 3.31), and coerces sufficiently these parameters to enable a slight increase (up to 20 %) of their standard-deviations (which also coerce the individual parameters to follow the pattern given by $f_{\mathbf{\Theta}|\mathbf{\Pi}}$). We can now continue to the analyses of the correlation coefficients. For both scenarios, the estimated correlations are similar. The low value of $\rho(E_0, d_c)$ suggests that this parameter is in fact 0 (hence its COV is not defined and labeled as non-attributed written N/A). The correlation between y_{cs} and d_c is close

to one, and consequently, the likelihood function is sensitive to it, explaining the low variability of its estimation (the associated coefficient of variation is smaller than 0.2%). The one between y_{cs} and E_0 is about half of $\rho(y_{cs}, d_c)$ (in absolute value), and therefore the likelihood function is less sensitive to it than $\rho(y_{cs}, d_c)$, explaining its greater variability (ten times bigger). The variability of the other population parameters is also limited, even for those that concern the damage threshold, but to a lesser extent.
These correlations can be explained from a physical point of view. Let's focus first on $\rho(E_0, y_{cs})$. Remember that y_{cs} indicates the speed of damage in the following way: the higher this parameter is, the slower damage increases and conversely. If we come back to Figure 3.38, all specimens reach the same level of damage (their saturation is similar). Thus, if the initial slope monitored by E_0 of one specimen is high, to reach the same level of damage, it has to develop faster, which is synonymous of a decreased y_{cs}. To explain the second dependency ($\rho(y_{cs}, d_c)$), we should remind first that if damage is slow, more cracks may emerge before failure. Provided that the damage saturation is related to the amount of cracks within the specimen, the latter sentence can be then "translated" in terms of model parameters as follows: high saturations imply high y_{cs} (which indicates the inverse of the damage speed). It is worth noticing, that, for the same model error, between the calibration scenarios, the differences come from different factors. For y_{cs} and d_c, it corresponds to the inclusion of the dependency modeling, while for E_0 it is the update of the marginal. This comes from the fact the dependency relation in which E_0 is involved is twice smaller in terms of magnitude than the one between y_{cs} and d_c.
The remaining coefficient to examine is the damage threshold y_{0s}. In the previous sections (3.5.4), we showed that the damage threshold was difficult to identify, especially because the acoustic data that should used to have a proper calibration of this parameter are not available in our test-case. Acoustic emission is indeed widely used to follow damage recording acoustic events. Table 3.29 also shows that for all experiments, the identification of this parameter depended more on the population parameters than for the others, suggesting that the calibrated values depends more on the regularization term than on the experimental data, highlighting a risk of error on the provided estimates. To assess the estimation of its marginal, this parameter appears as much more sensitive to the model error. This can be understood as follows: if we consider a single residual model for all experiments, the MLE estimate of ω, $\widehat{\omega}$, stands for the average of the function $g_i(\widehat{\boldsymbol{\theta}}_i, \mathbf{y}_i, \boldsymbol{\Psi})$ over all the experiments ([Pinheiro and Bates, 2002]). Thus, if the data from some specimens is flawed with more errors than the others, $\widehat{\omega}$ will be higher from what it would have been without consideration of the specimens that exhibit the highest uncertainty. The explanation is the result of the low sensitivity of the model output with respect to this parameter provided that ω influences the level of accuracy of the calibration (high ωs strongly penalize poor calibration of the model and conversely, see Eq. (3.20)). This also explains the higher E_0s: this parameter mainly expresses at the beginning of the stress-strain curve where the residuals are smaller compared to the other from the nonlinear behavior. With high ωs, the weight of the residuals from the linear behavior decreases which allows a less accurate estimation of the elastic part and thus of the individual modulus. To continue this evaluation, we can focus on the value taken by the damage thresholds. To do this, we can compute the strain damage threshold that is expressed as $\varepsilon_{0s} = \sqrt{\frac{2y_{0s}}{E_0}}$. A rough evaluation using the average values provides $1.23 \times 10^{-3}\%$ for scenario $\mathcal{D}_1$ and 0.0251% for the other scenarios. If we refer to Figure 3.38, given the strain values, it would mean that the specimens are already damaged at the start of the experiment. To get a proper analysis, we should rather analyze the individual thresholds. This is displayed in Figure 3.41.

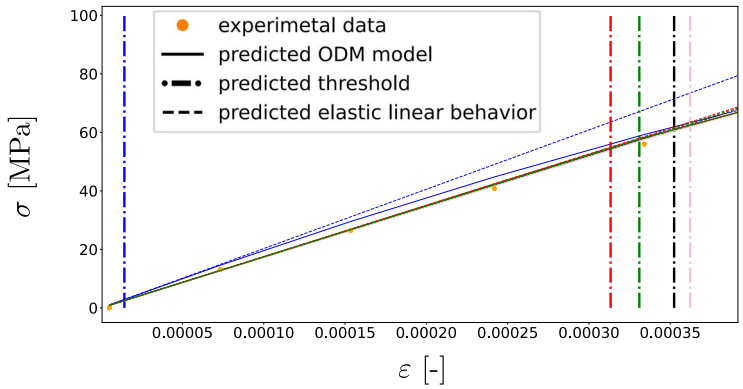


Figure 3.41: Zoom on the start of the stress-strain curve for the 2$^{\text{nd}}$ experiment. Model curves are colored according to the associated scenario. Only the individuals parameters from the best calibration (in term of likelihood) are employed provided their small variability (as discussed below). Model curves from all scenarios are superposed in the colors defined in Table 3.27.

To analyze Figure 3.41, we can start by noticing that the predicted responses of the elastic linear behavior and of the ODM model start to diverge after the beginning of damage, that asserts the consistency of the results for all scenarios. On this specimen (the 1$^{\text{st}}$), the first points of the stress-strain curves show to be more accurately approximated with a linear model than with a nonlinear behavior. In other words, the Young's modulus and damage thresholds estimated in scenarios $\mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4$ and $\mathcal{D}_5$ seem to be more consistent with the experimental data than those calibrated under scenario $\mathcal{D}_1$. In fact, if we retain the mathematical definition of the damage threshold, that is to say the end of the linear elastic behavior (or conversely the start of damage), the calibration of the thresholds from scenarios $\mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4$ and $\mathcal{D}_5$ provides better results than from $\mathcal{D}_1$. However, from a physical point of view, it is more difficult to draw clear conclusions. The damage threshold corresponds to the moment at which the first damage represented by the first cracks develops on the material [Ben Ramdane, 2014]. Physical studies have shown that cracks within the matrix and the fibers may occur during the manufacturing process [Ben Ramdane et al., 2017]. Thus, the material is damaged even before any stress is applied to the sample that justifies a strain threshold deformation smaller than the initial strain [Ben Ramdane et al., 2017, Débarre et al., 2022] as provided here. The calibrated copulas and marginals are respectively depicted in Figure 3.42 and 3.43.

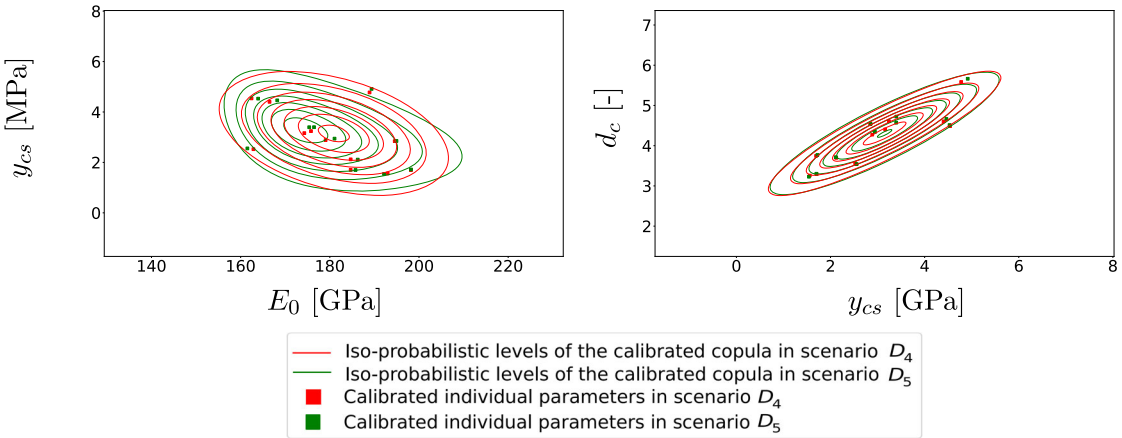


Figure 3.42: Calibrated dependency structure for scenarios $\mathcal{D}_4$ and $\mathcal{D}_5$ with 12 specimens from the experimental data. Individual parameters are also plotted. Only the copulas from the best calibration (in terms of likelihood) are drawn considering their small variabilities.

The calibration should also be evaluated at the individual level. As there is no exact value for the individual parameters in this case, the only available analysis consists in computing for all scenarios and specimens the RMSEs between the experimental curves and predicted models for

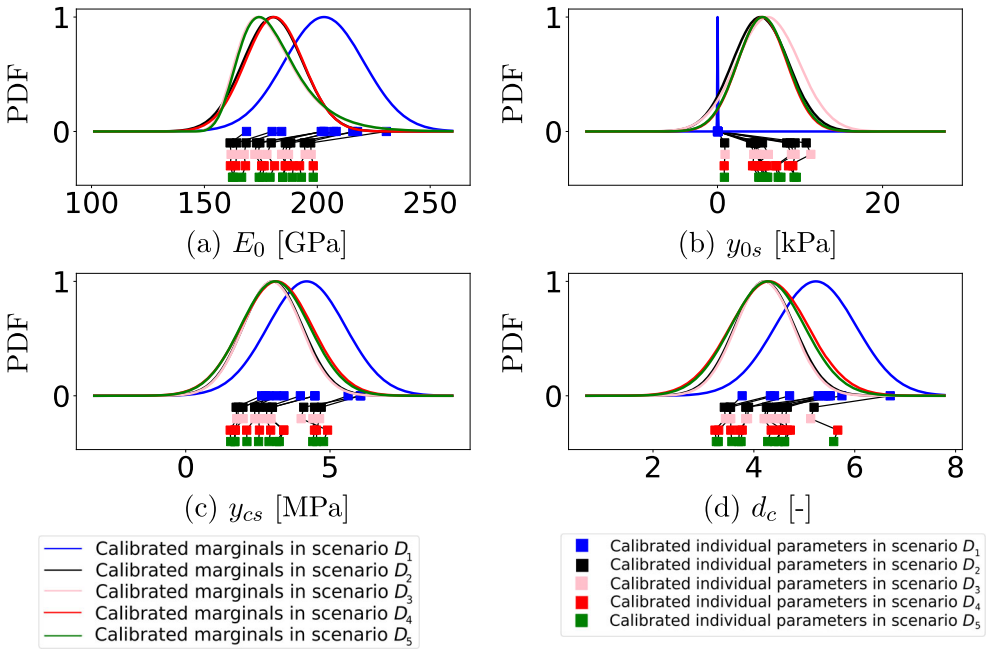


Figure 3.43: Calibrated marginals for all parameters with 12 specimens from the experimental data along the E_0-axis (a), y_{0s}-axis (b), y_{cs}-axis (c) and d_c-axis (d). Only the marginals from the best calibration (in terms of likelihood) are drawn considering their small variabilities (except for the damage threshold in scenario $\mathcal{D}_1$).

the estimated individual parameters. Furthermore, note that these RMSEs can be interpreted as a posterior estimation of the standard-deviation of the residuals which are supposed to correspond to the experimental noise as model bias is supposed to be negligible. Consequently, the estimated standard-deviation of the noise can be compared to the MLE estimate. These quantities are linked because the ω_is are tuned given the residuals values that result from the individual calibrations (Eq. (3.6)), where the ω_is, kept constant, influence the accuracy of the optimization, so indirectly the residuals. Those two are reported in Table 3.32 for experiment 12 which represents the two types of experiments available: those that break around a strain of 0.4% and those for which failure occurs later.

Table 3.32: Averaged calibrated and estimated standard-deviation of the residuals over 4 repetitions of the calibration process with 12 specimens each for experiments 2 and 12 in each scenario. The coefficients of variation COV in % are indicated between brackets.

	Experiment 2	Experiment 12
Averaged calibrated standard-deviation in scenario $\mathcal{D}_1$ [MPa]	2.88 (0.0840)	2.88 (0.0840)
Averaged estimated standard-deviation in scenario $\mathcal{D}_1$ [MPa]	1.92 (0.150)	3.09 (0.0876)
Averaged calibrated standard-deviation in scenario $\mathcal{D}_2$ [MPa]	1.51 (2.00)	3.78 (0.200)
Averaged estimated standard-deviation in scenario $\mathcal{D}_2$ [MPa]	1.42 (2.32)	3.50 (0.327)
Averaged calibrated standard-deviation in scenario $\mathcal{D}_3$ [MPa]	1.50 (0.105)	3.78 (0.0778)
Averaged estimated standard-deviation in scenario $\mathcal{D}_3$ [MPa]	1.39 (0.0833)	3.50 (0.0106)
Averaged calibrated standard-deviation in scenario $\mathcal{D}_4$ [MPa]	1.62 (0.210)	3.77 (0.110)
Averaged estimated standard-deviation in scenario $\mathcal{D}_4$ [MPa]	1.51 (0.0586)	3.48 (0.0609)
Averaged calibrated standard-deviation in scenario $\mathcal{D}_5$ [MPa]	1.59 (2.12)	3.78 (0.146)
Averaged estimated standard-deviation in scenario $\mathcal{D}_5$ [MPa]	1.47 (2.40)	3.50 (0.194)

From Table 3.32, we can first notice the accuracy of the calibration for each specimen: on average, for all specimens and scenarios, the RMSEs reach at most 4 MPa, indicating a precise prediction of the experimental stress that goes well over 100 MPa (so an error on the stress prediction smaller than 4%), regardless of the assumptions on the marginals, dependency and residual model, from which can be concluded that all specimens are properly identified. It also underlines the interest to consider different residual models for each specimen. Indeed, as we stated before, if the residual model is the same for all individuals, the determination of its standard-deviation is driven by the experiments with the highest errors and will thus be bigger than for the experiments that exhibit lesser measurement noises. This is illustrated here for the second experiment: the calibrated standard-deviation of the noise is much higher (50%) than the actual level of noise (the RMSE computed for the calibrated individual parameters), even if this difference drops for the 12^{th} experiment. This comes from the fact that it is the experiment with the highest RMSE, and it has a major impact on ω. On the contrary, with different residual models for each specimen, the estimated and calibrated standard-deviations of the noise demonstrate a better agreement with both experiments. We can note the small coefficient of variations of these two quantities, prompting that, for all repetitions of each scenario, roughly the same individual parameters are estimated. More precisely, it shows that the individual parameters calibrated with the same residual model error are similar, contrary to those identified with another residual model. Given that the estimation of the population parameters is driven by the individual parameters, it partly explains the difference between the population parameters for the different choices of residual model. For the same residual model, the differences between the $\widehat{\boldsymbol{\theta}}_i$s comes from the population modeling. The calibration of these two experiments are depicted in Figure 3.44.

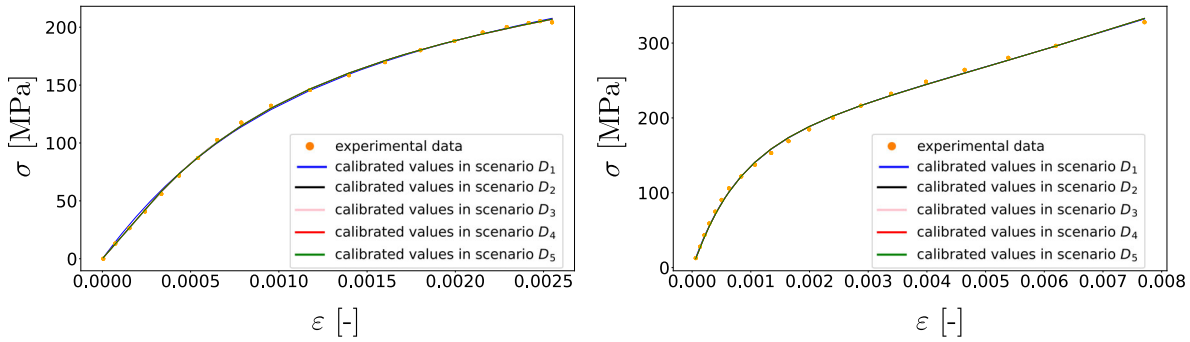


Figure 3.44: Calibrated models for experiments 2 (left) and 12 (right) for all scenarios.

A final check of the results is the analysis the residuals for the estimated individual parameters. They are displayed in Figure 3.45. It shows, as in Chapter 1, that within the residuals remain an oscillatory structure that may include among other sources model bias. Adding another term that would account for a model bias would represent a clear improvement of the method. This yet represents an increase of the method complexity, explaining why we keep in this work the same assumptions on the residual model. However, as a perspective, an additional bias model could be added to account for the structure of the residuals as a function of the strain that is here the model input. Otherwise, the residuals are symmetric, and are below 4 MPa for all observation points and scenarios, underlining the proper calibration of the experiments shown in Figure 3.45.

3.5.4 Assessing the calibrated distributions

Mixed-effects methods have been used to provide estimates of material variability. Probability distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ represents it in parameters space. Yet, to make structural forecasts (of failure strength for instance), we should estimate it in model space by propagating the different calibrated distributions through the ODM model. This is carried out by Monte-Carlo sampling (as described in Section 3.4.5). Then, for the 4 repetitions and the different scenarios, the upper

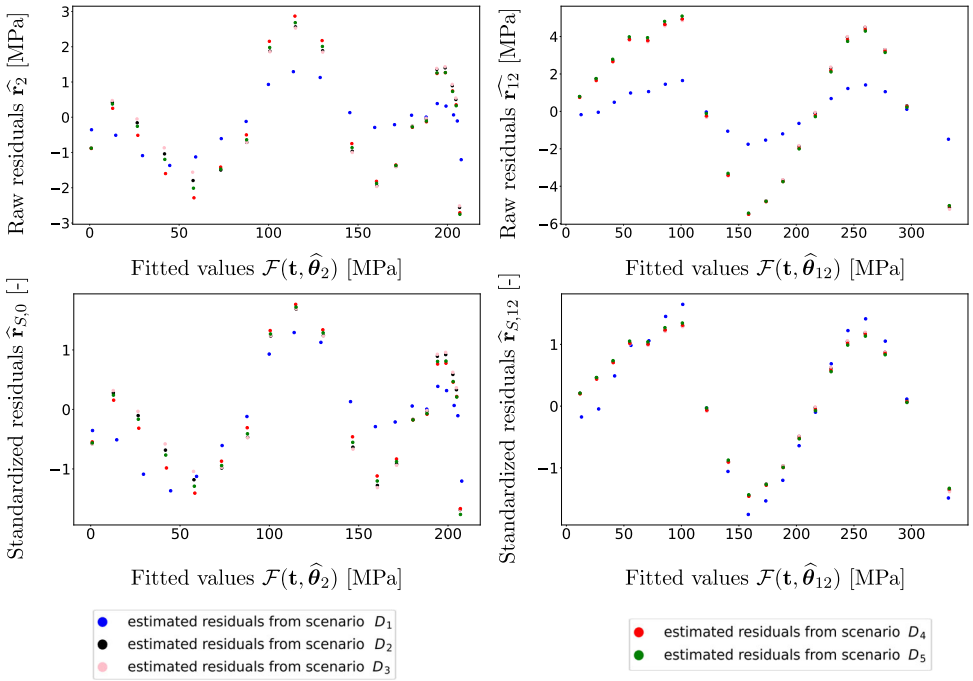


Figure 3.45: Estimated raw (on the top) and standardized residuals (on the bottom) for experiments 2 (on the left) and 12 (on the right) for all scenarios.

and lower quantiles (5% and 95%) are computed. The discrepancy between those two stand for the possible material behaviors, all proposed in Figure 3.46.

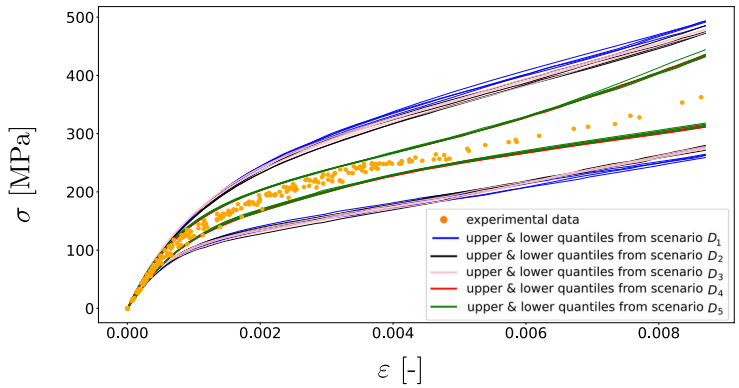


Figure 3.46: Propagation of the calibrated distributions for all scenarios and repetition of the calibration process.

The propagation of the calibrated distributions shows that for the scenarios including statistical dependencies, the estimated variability is more consistent with the experimental data. The corresponding upper and lower quantiles surround tightly the experimental data and widen for strain bigger than 0.4%, where data lacks because most of the specimens have been broken. This underlines the importance of the correlations in the modeling of material variability. In fact, this factor proves to be the most important, far ahead of the choice of the marginal between a Gaussian or a Gumbel marginal, or the impact of the residual error on the model parameters distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$. This can sound surprising because, in Table 3.30, the most influential one seemed to be the residual model (in terms on impact on the means and standard-deviations). In fact, here, as the experiments are suitably estimated with both residual models, its choice seems not to be a decisive feature. Here again, changing the marginals does not bring a specific improvement in the estimation of material variability, as depicts Figure 3.46. From the point of view of an engineer of a design office who aims to strike a balance between model complexity and accuracy, in this specific test-case, it appears that there is no need to use non-Gaussian

marginals and copulas. In other words, in this scope, the simplest modeling can be thought as the best one. If we have to choose amidst the different modeling assumptions, we would favor scenarios $\mathcal{D}_4$ and $\mathcal{D}_5$. Recall that the in the two scenarios that include correlations, three extra parameters are considered (so 3 extra degrees of freedom compared to scenarios $\mathcal{D}_2$ and $\mathcal{D}_3$ and even 16 compared to $\mathcal{D}_1$) and we may wonder whether this does not imply over-fitting. A relevant tool to investigate this aspect is the BIC criterion [Lavielle, 2014] already implemented in Section 3.4.5. Still, remember that the BIC is based on an asymptotic criterion and with only 13 available, conclusions should be drawn cautiously. The BICs for the different scenarios are detailed in Table 3.33

Table 3.33: Averaged BICs for all scenarios. The coefficients of variation COV in % are indicated between brackets.

	Scenario $\mathcal{D}_1$	Scenario $\mathcal{D}_2$	Scenario $\mathcal{D}_3$	Scenario $\mathcal{D}_4$	Scenario $\mathcal{D}_5$
Averaged BIC	-1876 (4.91×10^{-4})	-1873 (6.03×10^{-3})	-1872 (1.90×10^{-5})	-1893 (1.76×10^{-4})	-1894 (1.64×10^{-3})

From Table 3.33, the first conclusion that can be drawn is the modeling from scenarios $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$ is comparable, provided the number of available number of individuals. The modeling that gives the best compromise between estimation of the distribution and number of degrees of freedoms are scenarios $\mathcal{D}_4$ & $\mathcal{D}_5$. In fact, the key factor is here the dependency structure, changing the residual model and the marginals do not provide any specific improvement. However, the relative difference between the BICs from these two scenarios those from the others is smaller than 1%, which can be considered as negligible provided the small number of individuals.
This investigation can be repeated by adding the data processed by the method the last experiment. This enables to characterize the impact of an specimen for which experimental issues occurred on the calibration process.

3.5.5 Calibrating with the complete data

With the complete data, the same testing protocol is applied to the 13 experiments. The calibration is achieved with the same modeling assumptions as those detailed in Table 3.27. The new updated scenarios will be reported as $\mathcal{D}_q^c$, $q \in [\![1, 5]\!]$. Of course, as a new specimen is added to the database, in scenarios $\mathcal{D}_2^c$ and $\mathcal{D}_3^c$, 21 full population parameters have to estimated and 24 in scenarios $\mathcal{D}_4^c$ and $\mathcal{D}_5^c$. The new population parameters are detailed in Tables 3.34, 3.35 and 3.36.

Table 3.34: Averaged calibrated means $\widehat{\mathbf{\Pi}}_{\text{mean}}$ over 4 repetitions of the calibration process with 13 specimens each. The coefficients of variation COV in % are indicated between brackets.

Averaged calibrated means	E_0 [GPa]	y_{0s} [kPa]	y_{cs} [MPa]	d_c [-]
Scenario $\mathcal{D}_1^c$	202 (0.105)	0.120 (34.6)	4.22 (0.475)	5.23 (0.175)
Scenario $\mathcal{D}_2^c$	179 (0.230)	6.23 (4.13)	2.68 (6.85)	4.01 (2.90)
Scenario $\mathcal{D}_3^c$	179 (0.286)	6.01 (7.81)	2.70 (7.40)	4.03 (3.12)
Scenario $\mathcal{D}_4^c$	180 (0.224)	5.87 (0.361)	2.55 (0.150)	3.82 (0.0391)
Scenario $\mathcal{D}_5^c$	180 (0.0399)	5.47 (0.251)	2.53 (0.328)	3.81 (0.159)

Comparing Tables 3.34, 3.35 and 3.36 to Tables 3.30 and 3.31, the estimated population parameters have similar values (or trends for the correlations). Yet, the main difference between the estimations of the population parameters is the level of variability of calibrated values. This occurs for all parameters, in particular for the standard-deviations except $\mu_{E_0}, \text{sd}(E_0)$. For instance, from scenario $\mathcal{D}_2$ to scenario $\mathcal{D}_2^c$, the COV of $\mu_{y_{cs}}$ increases from 1.06% to 6.85%, the COV of $\text{sd}(d_c)$ increases from 0.663% to 21.8%. However, no specific increase is spotted for

Table 3.35: Averaged calibrated standard-deviations $\widehat{\mathbf{\Pi}}_{\text{mean}}$ over 4 repetitions of the calibration process with 13 specimens each. The coefficients of variation COV in % are indicated between brackets.

Averaged calibrated standard-deviations	E_0 [GPa]	y_{0s} [kPa]	y_{cs} [MPa]	d_c [-]
Scenario $\mathcal{D}_1^c$	17.5 (0.464)	0.318 (27.3)	1.32 (0.533)	0.800 (0.674)
Scenario $\mathcal{D}_2^c$	12.5 (1.95)	3.59 (3.93)	1.26 (14.8)	0.882 (21.8)
Scenario $\mathcal{D}_3^c$	14.4 (0.783)	3.58 (4.40)	1.27 (14.9)	0.881 (22.5)
Scenario $\mathcal{D}_4^c$	12.6 (0.262)	2.85 (0.440)	1.43 (0.493)	1.08 (0.415)
Scenario $\mathcal{D}_5^c$	14.6 (0.416)	2.87 (0.487)	1.40 (0.468)	1.08(0.246)

Table 3.36: Averaged calibrated correlations $\widehat{\mathbf{\Pi}}_{\text{mean}}$ over 4 repetitions of the calibration process with 13 specimens each. The coefficients of variation COV in % are indicated between brackets. N/A stands for non-attributed.

Averaged calibrated correlations	$\rho(E_0, y_{cs})$	$\rho(E_0, d_c)$	$\rho(y_{cs}, d_c)$
Scenario $\mathcal{D}_4^c$	-0.240 (1.53)	0.09 (N/A)	0.916 (0.0584)
Scenario $\mathcal{D}_5^c$	-0.285 (1.10)	0.05 (N/A)	0.911 (0.0521)

the scenarios in which correlations are included. One of the reason why it behaves as such is that, as stated earlier, on the last experiment (the 13$^{\text{th}}$), the estimated individual parameters depend more on the population parameters than for the other experiments, because the failure occurs too early to enable a proper calibration of the damage saturation and of the damage evolution celerity. It should be underlined that this phenomenon is not unexpected and comes numerically from the fact that the fitting term $(-\ln(f(\mathbf{y}_i|\boldsymbol{\theta}_i, \mathbf{\Pi}, \omega_i)))$ is of the same order of magnitude than the regularization term $(-\ln(f(\boldsymbol{\theta}_i|\mathbf{\Pi})))$ contrary to the other experiments (see Table 3.29 and Figure 3.40). This does not occur when statistical dependencies are included which can be explained by the fact that statistical dependencies coerce individual parameters, therefore decreasing the variability of their estimates, and consequently, the same applies to the variability of the population parameters estimates.Still, this point would require to be inquired further. The variability on the determination of the individual parameters implies an increase of the variability of the population parameters, as illustrated in the previous Tables, especially given the small number of available experiments. In fact, this is similar to what happens to the damage threshold, but with greater consequences given the sensitivity of the model output to these parameters. This can also be noticed for the standard-deviations of the residual model, $\omega_{(i)}$ (see Tables 3.37 and 3.38).

Table 3.37: Averaged calibrated and estimated standard-deviation of the residual model $\mathbf{\Psi}_{\text{mean}}$ over 4 repetitions of the calibration process with 13 specimens each for experiment 12 and 13 in scenarios $\mathcal{D}_1^c, \mathcal{D}_2^c$ and $\mathcal{D}_3^c$ (data for other scenarios is next page). The coefficients of variation COV in % are indicated between brackets.

	Experiment 12	Experiment 13
Averaged calibrated standard-deviation in scenario $\mathcal{D}_1^c$ [MPa]	2.79 (0.0860)	2.79 (0.0860)
Averaged estimated standard-deviation in scenario $\mathcal{D}_1^c$ [MPa]	3.09 (0.0748)	1.20 (0.0153)
Averaged calibrated standard-deviation in scenario $\mathcal{D}_2^c$ [MPa]	3.84 (1.07)	0.836 (26.6)
Averaged estimated standard-deviation in scenario $\mathcal{D}_2^c$ [MPa]	3.54 (0.340)	0.759 (28.0)
Averaged calibrated standard-deviation in scenario $\mathcal{D}_3^c$ [MPa]	3.75 (0.200)	0.837 (26.3)
Averaged estimated standard-deviation in scenario $\mathcal{D}_3^c$ [MPa]	3.50 (0.233)	0.762 (28.0)

Table 3.38: Averaged calibrated and estimated standard-deviation of the residual model $\boldsymbol{\Psi}_{\text{mean}}$ over 4 repetitions of the calibration process with 13 specimens each for experiment 12 and 13 in scenarios $\mathcal{D}_4^c$ and $\mathcal{D}_5^c$ (continues previous table). The coefficients of variation COV in % are indicated between brackets.

	Experiment 12	Experiment 13
Averaged calibrated standard-deviation in scenario $\mathcal{D}_4^c$ [MPa]	3.79 (0.130)	0.700 (0.202)
Averaged estimated standard-deviation in scenario $\mathcal{D}_4^c$ [MPa]	3.51 (0.130)	0.633 (0.917)
Averaged calibrated standard-deviation in scenario $\mathcal{D}_5^c$ [MPa]	3.77 (0.492)	0.700 (0.478)
Averaged estimated standard-deviation in scenario $\mathcal{D}_5^c$ [MPa]	3.51 (0.0285)	0.633 (0.204)

In addition to the higher variability of both the estimated and calibrated standard-deviations of the noise (except for scenario $\mathcal{D}_1^c$, $\mathcal{D}_4^c$ and $\mathcal{D}_5^c$), Tables 3.37 and 3.38 also illustrate that adding the 13$^{\text{th}}$ experiment to the other 12 does not alter the identification of their respective individual parameters. This is a key feature, consistent with the observations of Table 3.29, that indicate that the estimation of the individual parameters for all other specimens was rather independent from the chosen distribution. In fact, the scenarios for which the population parameters exhibit large variability match with those where the same applies to the $\widehat{\omega}_{(i)}$s and $\boldsymbol{\theta}_i$s. For scenario $\mathcal{D}_1^c$, it can be explained by the choice of residual model. Indeed, the 13$^{\text{th}}$ is the one with the smallest RMSE. Thus, the estimated standard-deviation of the residual model remains influenced by those with the highest noise, imposing the same level of accuracy as we had without it and consequently, an identical identification of the individual parameters and the population parameters prevail. This variability can also be illustrated by propagating the estimated distribution as performed in Figure 3.47.

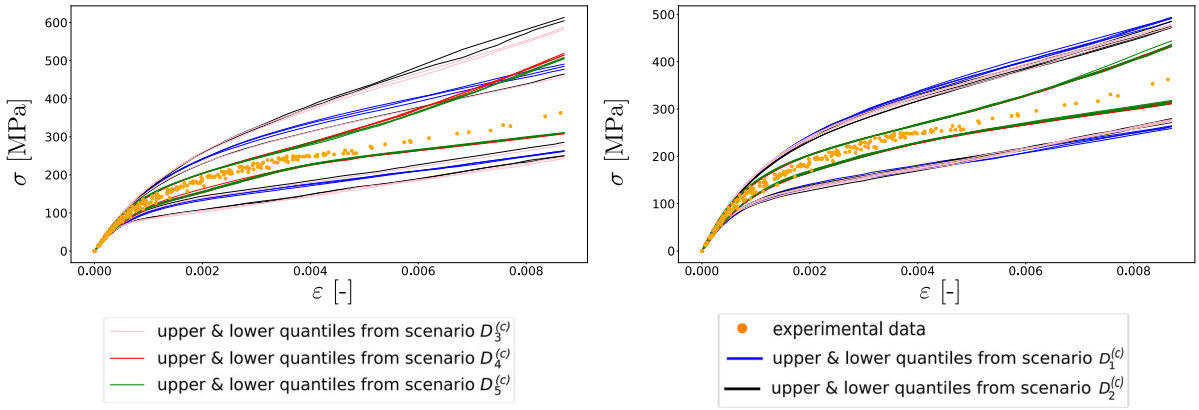


Figure 3.47: Propagation of the calibrated distributions for all repetitions and scenarios with all specimens (on the left) and with only 12 (on the right).

In Figure 3.47, the consequences of the variability of the calibration of $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ are illustrated. Comparing both pictures, we can see that adding one experiment to the database does not improve but rather downgrades its estimation, as it widens the set of possible outputs to behavior that are not consistent with the available experimental data. It shows that, with few individuals, it is important to check that the model parameters can be identified suitably on all experiments, otherwise, the estimation of $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ may get flawed. This effect decreases as the number of experiments (lead without issues) increases. Still, the method presented provides reasonable estimate of material variability that can be used for further analyses. Among the possible improvement lies the enhancement of the residual model by adding a bias term and to account for multiple sources of information, in this case the combination of the stress-strain curve and acoustic data. Finally, to consolidate the results, the uncertainty on the population parameters should be carried out to have proper prediction.

3.6 Conclusion

In this chapter, we proposed a first implementation of mixed-effects in the context of a material behavior law calibration to investigate the different settings of the method, *e.g.*, the influence of the number of individuals, the dependency modeling and the choice of the marginals. Our results demonstrate that for various situations, mixed-effects methods allow to estimate properly both the population and distribution parameters. Three major conclusions arise. First, the number of available experiments is a key feature. Indeed, increasing the number of individuals decreases both the estimation error and the uncertainty (represented by the size of the 95% confidence interval) on the population parameters. Furthermore, except in very specific cases, the individuals (either in model and individual spaces) are most of the time suitably calibrated in terms of calibration errors following the limitations coming from the low sensitivity of the damage threshold and the number of available individuals. The second conclusion that can be drawn from our observations is that, with few individuals, the estimation of the full population parameters (and in particular of the dependency structure) can be significantly altered, but this phenomenon declines with the number of individuals. Still, with few individuals, because of lack of specimens, the calibrated distribution with the exact assumptions (that is to say when statistical dependencies are added to $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$) does not necessarily provide better estimations of the exact model parameters distribution than when all parameters distribution, even if this is in this case a false assumption. In addition, with very few individuals (that is to say about 10), a convenient identification of the individual parameters, a key feature in all cases, becomes decisive because if any of the individual is not properly calibrated, the whole estimation can be either seriously downgraded or made uncertain, even if this risk also diminishes with the number of individuals. Especially, if the sensitivity of the model output with respect to one of the model parameters is low (as we saw with the damage threshold), the corresponding marginal will be poorly estimated, the same applying for the dependency structure. To identify properly a given marginal, it is necessary for the corresponding parameter to be suitably calibrated on most (if not all) experiments, and the same applies for the statistical dependencies. Another remarkable point that can be highlighted is in the impact of the dependency modeling and marginals on the propagated material variability through the model. It appears that the shape of the distribution has an impact, but at a second-level, contrary to the dependency that significantly impacts the estimation of the material variability. However, the study presented in this work was limited to distribution whose PDFs share similar smoothness properties and shapes. With marginals (and copulas) that are significantly more different, the conclusion could have been dramatically different. The study with experimental data has confirmed these previous observations. The method seems to provide the acceptable characterization of material variability that showed to be consistent with the available data and the physical knowledge on the behavior of such materials.

This first application of mixed-effects on a simplified uniaxial nonlinear model has proven to be efficient to provide an estimation of material variability. Key features such as the number of individuals, the modeling of dependency structure and the sensitivity of the model output with respect to the model parameters have been highlighted. These topics have been investigated on synthetic data that has helped our analysis of the mixed-effects calibration of the ODM model results with real data. Yet, even if it is more complicated than most of the pharmocokinetic models used in the literature related to mixed-effects [Lavielle, 2014], this one remains far more easy to handle than behavior laws used in practice as they involve multi-axial (*e.g.*, longitudinal and transverse) or multi-scale behaviors [Defense et al., 1999], implying a significant number of model parameters but also different sensitivities of the model output with respect to the model parameters on the usual experiments. More specifically, not all parameters express on the same experiment. For instance, longitudinal parameters do not influence the transverse behavior and conversely [Berthelot, 2012]. Then, from our conclusions of this chapter, it follows that not all population parameters can be estimated at once from the same experiment. However, if the

current modeling of the individual parameters defined in Eq. (3.2) is extended, depending on the model to calibrate, standard mixed-effects allow to calibrate all population parameters at once. Still, this imply to estimate distributions in high dimensions (up to 20 for the complete ODM model), making the calibration difficult if not intractable to carry out. Another solution that aims to estimate in multiple steps the population parameters can be also set up. We choose the second solution and propose in the next chapter a sequential procedure.

CHAPTER 4

ADAPTATION OF MIXED-EFFECTS TO A SEQUENTIAL APPROACH

Contents

Chapter contributions:

- Adaptation of the mixed-effects model framework to the use of multivariate models and data
- Proposition of a sequential strategy to handle the calibration of complex behavior laws with mixed-effects
- Implementation of the sequential strategy for the calibration of a nonlinear elastic model with synthetic and experimental data
- Propagation of the calibrated distributions to estimate a failure stress to assess the validity of the calibration assumptions

4.1 Introduction

The production of composite materials is subjected to many sources of uncertainty that impact the material properties [Defense et al., 1999]. To quantify this variability, mixed-effects methods that belong to the population approaches described in Chapter 2 have been implemented in Chapter 3 with a uniaxial simplified version of the ODM model. We demonstrated the ability of mixed-effects to provide consistent estimates of material variability contrary to the techniques studied in Chapter 1. Several settings of the method such as the number of available individuals have been examined. However, the examples studied in the previous chapters have been made on quite simple problems. More advanced models used to predict the performance of composite materials, such as the OPFM model [Laurin, 2005], induce more complex behavior laws to be estimated. These behavior laws involve the combination of several phenomena (elasticity, failure, damage, viscoelasticity *etc.*) to which are attached specific parameters. They also describe in-plane and out-of-plane behaviors (multi-axial model). Thus, the model to calibrate usually involves a large number of parameters, increasing the size of the search space and making the optimization of the likelihood more difficult. The inherent heterogeneity and anisotropy of composites make the number of model parameters even larger. In fact, to calibrate the material parameters, experimental tests of different natures are often requested, either in terms of loads (*e.g.*, tension, compression or shear tests) or in terms of stacking sequences (*e.g.*, unidirectional $0°$ or quasi-isotropic laminates). Nevertheless, the sensitivity of the model output with respect to the model parameters depends on the nature of the test [Berthelot, 2012] and all the model parameters characterizing a given phenomenon (elasticity for instance) cannot always be estimated with a unique type of test. Consequently, it could be interesting to consider a calibration strategy based on a decomposition of the behavior law according to the type of described phenomenon. This would reduce the number of model parameters to be estimated, and consequently the number of population parameters to calibrate. Such a process could decrease the complexity of the calibration in the presence of uncertainty by taking advantage of the decomposition of the model by transforming the overall calibration problem into a sequence of appropriate smaller calibration subproblems, corresponding to the different test configurations (*e.g.*, different load profiles and stacking sequences). This calibration method reproduces the engineer's approach for model identification that consists in estimating the model parameters on several experiments.

These are designed to isolate the influence of each parameter to ease its estimation. In addition to decreasing the number of parameters to be identified, this alleviates the computational costs as it avoids to take into account all the phenomena at the same time, which is a key feature with mixed-effects given that the computation of the associated likelihood (the fitting criterion) is challenging as it is defined as a product of multivariate integrals. This does not prevent to implement complementary strategies (*e.g.*, meta-modeling [Forrester et al., 2008]) to speed up the complete process.
The objective of this chapter is to introduce a sequential procedure to calibrate material behavior laws compliant with the mixed-effects framework. This procedure aims to calibrate a multi-axial orthotropic elastic model on multiple experiments with different stacking sequences. Once the method is numerically validated, the ultimate goal is to process with this sequential method experimental data coming from repetitions of tensile tests achieved on a composite material with carbon fibers and epoxy matrix. Finally, uncertainty propagation is performed in order to assess the validity of the calibration process. This chapter is organized as follows. First, a brief recap of the population approaches and mixed-effects is presented in Section 4.2 before detailing the treatment of multivariate data with mixed-effects. In Section 4.3, the proposed sequential calibration process is presented in details. In order to illustrate the performance of the calibration strategy, the methodology is applied to the calibration of an orthotropic elastic model with nonlinear longitudinal behavior. The test-case is described in Section 4.4 with a focus on the behavior law and the implementation of the sequential strategy. A numerical study is first carried out to investigate the performance of this methodology with synthetic data in Section 4.5, before being applied to experimental data from laminates made of a unidirectional carbon/epoxy base ply[1]. In Section 4.7, the calibrated distributions are propagated to estimate the variability of the failure stress of an open-hole laminated plate in tension, before concluding in Section 4.8. The first part of this chapter (up to Section 4.5.5 included) is based on a paper published in Mechanics & Industry: [Laboulfie et al., 2023].

4.2 Mixed-effects for multivariate models

The calibration of mixed-effects for univariate models has been extensively discussed [Demidenko, 2013, Lavielle, 2014, Pinheiro and Bates, 2002]. This happens when only one measure is available on each individual. Yet, sometimes, one has access to several measures on the same individual (with longitudinal and transverse gauges or digital image correlation for instance). Then, the model to be calibrated becomes multivariate in the sense that the output is no longer a scalar but rather a vector, requiring to adapt the mixed-effects equations [Hall and Clutter, 2004] and consequently, the corresponding likelihood. Section 4.2.1 recaps the statistical framework for univariate mixed-effects and its adaptation to handle multivariate data. The changes on the likelihood formulation and estimation are described in Section 4.2.2.

4.2.1 Adapting the statistical framework to multivariate data

As stated in the Chapter 2, the population aims to characterize the variability that arises through repetitions of experiments and requires to model both the population and individual levels. Let's concentrate on the case of univariate measures such as those used in Chapter 3. Let's note n the number of available repetitions with i the corresponding index, N_i the number the observations of the i^{th} subject with j the corresponding index of the j^{th} observation. The set of model parameters that characterizes the i^{th} specimen representing the individual parameters is labeled $\boldsymbol{\theta}_i$ and of dimension d. The j^{th} input and output data of the i^{th} specimen are respectively labeled t_{ij} and y_{ij}. The behavior law is denoted $\mathcal{F}(\cdot,\cdot)$. The residual term is labeled ξ_{ij}. The probability distribution that represents material variability, called the model parameters

[1]The material is not named for confidentiality issues

distribution, is labeled $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ ($\boldsymbol{\Pi}$ stands for to the population parameters, that is to say the hyperparameters of the model parameters probability distribution and is of dimension s). The equations governing the mixed-effects are [Demidenko, 2013, Lavielle, 2014, Pinheiro and Bates, 2002]:

$$\begin{cases} \forall i \in [\![1,n]\!], \boldsymbol{\theta}_i \underset{\text{i.i.d.}}{\sim} f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}} \ , & \text{(4.1a)} \\ \forall (i,j) \in [\![1,n]\!] \times [\![1,N_i]\!] \ , y_{ij} = \mathcal{F}(t_{ij}, \boldsymbol{\theta}_i) + \xi_{ij} \ , & \text{(4.1b)} \\ \forall (i,j) \in [\![1,n]\!] \times [\![1,N_i]\!] \ , \xi_{ij} \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \omega_{ij}^2) \ , & \text{(4.1c)} \end{cases}$$

with ω_{ij} the standard-deviation of the noise of the j^{th} output measure of the i^{th} individual. Within the mixed-effects framework, the individual parameters are decomposed between fixed-effects $\boldsymbol{\beta}$ and random-effects $\mathbf{b}_i$ as:

$$\boldsymbol{\theta}_i = \boldsymbol{\beta} + \mathbf{b}_i \ . \tag{4.2}$$

Eq. (4.2) can be generalized by adding design matrices $\mathbf{A}_i$ and $\mathbf{B}_i$ into

$$\boldsymbol{\theta}_i = \mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \mathbf{b}_i \ . \tag{4.3}$$

Such matrices can be used for instance when different types of experiments are considered together.

To present the adaptation of mixed-effects to multivariate data, let's first introduce m_i the number of output measures of the i^{th} individual and $q \in [\![1, m_i]\!]$ the corresponding index. In the following, all individuals are supposed to have the same number of output measures, so $m_i = m, \ \forall i \in [\![1,n]\!]$. Moreover, all measures are supposed to have the same number of observation points N_i. The main difference between Eq. (4.1b) and Eq. (4.4b) is that the model output of each observation point is no longer a scalar y_{ij} but a vector $\mathbf{y}_{ij} \in \mathbb{R}^m$. The same occurs for the discrepancy term, now labeled $\boldsymbol{\xi}_{ij}$. The equations governing multivariate mixed-effects are [Demidenko, 2013, Lavielle, 2014, Pinheiro and Bates, 2002]:

$$\begin{cases} \forall i \in [\![1,n]\!], \boldsymbol{\theta}_i \underset{\text{i.i.d.}}{\sim} f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}} \ , & \text{(4.4a)} \\ \forall (i,j) \in [\![1,n]\!] \times [\![1,N_i]\!] \ , \mathbf{y}_{ij} = \mathcal{F}(t_{ij}, \boldsymbol{\theta}_i) + \boldsymbol{\xi}_{ij} \ , & \text{(4.4b)} \\ \forall (i,j) \in [\![1,n]\!] \times [\![1,N_i]\!] \ , \boldsymbol{\xi}_{ij} \underset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_{\mathbb{R}^m}, \boldsymbol{\Omega}_{ij}) \ , & \text{(4.4c)} \end{cases}$$

with $\mathbf{0}_{\mathbb{R}^m}$ the zero vector of $\mathbb{R}^m$ and $\boldsymbol{\Omega}_{ij}$ the covariance matrix of the discrepancy term which describes both the errors of the different sensors and the possible correlations between the errors on two different sensors. The observations are supposed serially uncorrelated as in earlier studies [McFarland et al., 2008], meaning that the measurement errors at two different observation times are independent (because uncorrelated Gaussian random vectors are independent):

$$\forall i \in [\![1,n]\!], \ \forall (j_1, j_2) \in [\![1, N_i]\!]^2, \ j_1 \neq j_2 \Rightarrow \text{Cov}(\boldsymbol{\xi}_{ij_1}, \boldsymbol{\xi}_{ij_2}) = 0 \ . \tag{4.5}$$

Furthermore, for all individuals, the noise is supposed to be homoscedastic, that is to say : $\boldsymbol{\Omega}_{ij} = \boldsymbol{\Omega}_i \ \forall \ j \in [\![1, N_i]\!]$. Note that the only difference between Eqs. (4.1b) and (4.4b) is that all terms are now vectors rather than scalars and that Eq. (4.4c) is the analogous of Eq. (4.1c) for vectors instead of scalars.

Let's define for all individuals $\mathbf{Y}_i$, the matrix of the output data of the i^{th} individual with $\mathbf{Y}_i := (\mathbf{y}_{i1}, \ldots, \mathbf{y}_{iN_i}) \in \mathcal{M}_{mN_i}(\mathbb{R})$. The rows of $\mathbf{Y}_i$ correspond to $\mathbf{y}_i^q = (y_{ij}^q)_{j \in [\![1,N_i]\!]}$, that is to say the N_i observations of the q^{th} measure of the i^{th} individual. Similarly, it is possible to define the matrix of the errors as $\boldsymbol{\xi}_i := (\boldsymbol{\xi}_{i1}, \ldots, \boldsymbol{\xi}_{iN_i}) \in \mathcal{M}_{mN_i}(\mathbb{R})$. The rows of the $\boldsymbol{\xi}_i$ matrix

correspond to $\boldsymbol{\xi}_i^q = (\xi_{ij}^q)_{j\in[\![1,N_i]\!]}$ and can be understood as the N_i measurement errors of the q^{th} measure for the i^{th} individual. Let's define $\widetilde{\mathbf{Y}}_i$ the flattened array by row of the matrix of the output data $\mathbf{Y}_i$ and the same for the matrix of the discrepancy term $\widetilde{\boldsymbol{\xi}}_i$. Then, $\widetilde{\mathbf{Y}}_i$ and $\widetilde{\boldsymbol{\xi}}_i$ belong to $\mathbb{R}^{mN_i}$. This allows to transform Eq. (4.4) into

$$\begin{cases} \forall i \in [\![1,n]\!], \boldsymbol{\theta}_i \underset{\text{i.i.d.}}{\sim} f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}} \ , & \text{(4.6a)} \\ \forall i \in [\![1,n]\!], \widetilde{\mathbf{Y}}_i = \mathcal{F}(\mathbf{t}_i, \boldsymbol{\theta}_i) + \widetilde{\boldsymbol{\xi}}_i \ , & \text{(4.6b)} \\ \forall i \in [\![1,n]\!], \ \widetilde{\boldsymbol{\xi}}_i \underset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_{\mathbb{R}^{mN_i}}, \widetilde{\boldsymbol{\Omega}}_i) \ , & \text{(4.6c)} \end{cases}$$

with $\widetilde{\boldsymbol{\Omega}}_i = \boldsymbol{\Omega}_i \otimes \mathbf{I}_{N_i}$ where $\otimes$ denotes the Kronecker product, $\mathbf{I}_{N_i}$ the identity matrix of $\mathcal{M}_{N_i}(\mathbb{R})$ and $\mathbf{0}_{\mathbb{R}^{mN_i}}$ the zero vector of $\mathbb{R}^{mN_i}$. Hence, the set of the model parameters of dimension p to be calibrated, labeled $\boldsymbol{\Psi}$ is defined as:

$$\boldsymbol{\Psi} \coloneqq \begin{bmatrix} \boldsymbol{\mu}, & \boldsymbol{\Sigma}, & \boldsymbol{\Omega}_1, & \ldots, & \boldsymbol{\Omega}_m \end{bmatrix} \ .$$

4.2.2 Likelihood estimation with multivariate data

As detailed in Section 2.3.2.2 of Chapter 2, the likelihood for mixed-effects with univariate data refers to $f(\mathbf{y}_1, \ldots, \mathbf{y}_n|\boldsymbol{\Psi})$, with f a generic letter for probability density functions (PDFs). For mixed-effects with multivariate data, the likelihood simply becomes $f(\widetilde{\mathbf{Y}}_1, \ldots, \widetilde{\mathbf{Y}}_n|\boldsymbol{\Psi})$. In fact, it remains the same except for $\mathbf{y}_i$ that becomes $\widetilde{\mathbf{Y}}_i$. The likelihood $\mathcal{L}(\boldsymbol{\Psi})$, which expresses as

$$\mathcal{L}(\boldsymbol{\Psi}) = f(\widetilde{\mathbf{Y}}_1, \ldots, \widetilde{\mathbf{Y}}_n|\boldsymbol{\Psi}) = \prod_{i=1}^{n} \mathcal{L}_i(\boldsymbol{\Psi}) = \prod_{i=1}^{n} \int_{\mathbb{R}^d} f(\widetilde{\mathbf{Y}}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \widetilde{\boldsymbol{\Omega}}_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) \mathrm{d}\boldsymbol{\theta}_i \ , \tag{4.7}$$

is challenging to compute because of the presence of the multivariate integrals. To estimate the likelihood, the Laplace approximation is used, given its ability to provide relevant estimations of the population parameters. This demands to rewrite the individual likelihoods $\mathcal{L}_i(\cdot)$:

$$\mathcal{L}_i(\boldsymbol{\Psi}) = \int_{\mathbb{R}^d} f(\widetilde{\mathbf{Y}}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \widetilde{\boldsymbol{\Omega}}_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) \mathrm{d}\boldsymbol{\theta}_i = \int_{\mathbb{R}^d} \exp(-g_i(\widetilde{\mathbf{Y}}_i, \boldsymbol{\theta}_i, \boldsymbol{\Psi})) \mathrm{d}\boldsymbol{\theta}_i \ , \tag{4.8}$$

with function $g_i(\cdot)$ defined by:

$$\begin{aligned} g_i(\widetilde{\mathbf{Y}}_i, \boldsymbol{\theta}_i, \boldsymbol{\Psi}) &\coloneqq -\ln(f(\widetilde{\mathbf{Y}}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \widetilde{\boldsymbol{\Omega}}_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})) = -\ln(f(\widetilde{\mathbf{Y}}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \widetilde{\boldsymbol{\Omega}}_i)) - \ln(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})) \\ &= \frac{1}{2}\|\boldsymbol{\Lambda}_i(\widetilde{\mathbf{Y}}_i - \mathcal{F}(\mathbf{t}_i, \boldsymbol{\theta}_i))\|_2^2 + \sum_{i=1}^{n} \frac{1}{2}\ln(|\widetilde{\boldsymbol{\Omega}}_i|(2\pi)^{mN_i}) - \ln(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})) \ , \end{aligned} \tag{4.9}$$

with $\boldsymbol{\Lambda}_i$ is the transpose of the Cholesky decomposition of $\widetilde{\boldsymbol{\Omega}}_i^{-1}$ (so $\widetilde{\boldsymbol{\Omega}}_i^{-1} = \boldsymbol{\Lambda}_i^\top \boldsymbol{\Lambda}_i$). In Eq. (4.9), the term $-\ln(f(\widetilde{\mathbf{Y}}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \widetilde{\boldsymbol{\Omega}}_i))$ is expanded following Eq. (4.6c). Applying formulas presented in Chapter 2, the Laplace approximation of the likelihood stands for

$$\begin{aligned} -\ln(\mathcal{L}(\boldsymbol{\Psi})) &= -\sum_{i=1}^{n} \ln\left(\mathcal{L}_i(\boldsymbol{\Psi})\right) = -\sum_{i=1}^{n} \ln\left(\int_{\mathbb{R}^d} f(\widetilde{\mathbf{Y}}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \widetilde{\boldsymbol{\Omega}}_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) \mathrm{d}\boldsymbol{\theta}_i\right) \\ &\approx \sum_{i=1}^{n} \left(\frac{1}{2}\ln(|\mathcal{H}(g_i)(\widehat{\boldsymbol{\theta}}_i)|) + \frac{g_i(\widetilde{\mathbf{Y}}_i, \widehat{\boldsymbol{\theta}}_i, \boldsymbol{\Psi})}{2}\right) - \frac{d}{2}\ln(2\pi) \ . \end{aligned} \tag{4.10}$$

In Eq. (4.10), $\widehat{\boldsymbol{\theta}}_i$ refers to the solution of the following minimization problem:

$$\widehat{\boldsymbol{\theta}}_i = \underset{\boldsymbol{\theta}_i \in \Gamma_{\boldsymbol{\theta}}}{\arg\min}\, g_i(\widetilde{\mathbf{Y}}_i, \boldsymbol{\theta}_i, \boldsymbol{\Psi}) \ , \tag{4.11}$$

with $\Gamma_{\boldsymbol{\theta}}$ the set of admissible parameters representing physical and mathematical constraints. In Eq. (4.10), $\mathcal{H}(g_i)(\cdot)$ is the Hessian of the function $g_i(\cdot)$ defined in Eq. (4.11) and can be derived as [Pinheiro and Bates, 2002]:

$$\mathcal{H}(g_i)(\widehat{\boldsymbol{\theta}}_i) \approx \frac{\partial \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i}^\top \boldsymbol{\Lambda}_i^\top \boldsymbol{\Lambda}_i \frac{\partial \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i} - \frac{\partial^2 \ln(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\widehat{\boldsymbol{\theta}}_i|\boldsymbol{\Pi}))}{\partial \boldsymbol{\theta}_i \boldsymbol{\theta}_i^\top} \,. \tag{4.12}$$

For Gaussian distributed individual parameters, (that is to say $\boldsymbol{\theta}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}$ the mean vector and $\boldsymbol{\Sigma}$ the covariance matrix), $\frac{\partial^2 \ln(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\widehat{\boldsymbol{\theta}}_i|\boldsymbol{\Pi}))}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_i^\top}$ reduces to $-\boldsymbol{\Sigma}^{-1}$.

From a practical point of view, the implementation of the code has been carried out in Python with the NumPy library [Harris et al., 2020] for matrix calculation and OpenTURNS library [Baudin et al., 2016] for the probabilistic modeling. It follows the same logic as the one presented in the appendices A

4.3 Sequential calibration process

Advanced mechanical behavior laws characterize different aspects of the mechanical phenomena of the material they describe, among which elasticity, viscoelasticity behavior, damage, *etc.* Each of these phenomena may be described by its own set of model parameters, which finally makes the number of parameters to be calibrated large (21 for the $3-d$ anisotropic elastic model for instance [Berthelot, 2012]). Consequently, a common practice [Schwer, 2006, Oberkampf and Roy, 2010] in classical deterministic calibration consists in dividing the full calibration problem into a sequence of calibration subproblems that enable to estimate different subsets of the complete parameters vector [Jung et al., 2022, Youn et al., 2011]. In these works, the different subproblems are associated with a specific partition of the model parameters. The method consists in solving a sequence of calibration problems taking advantage in each step of the knowledge acquired so far. To be more precise, for a given step, in a frequentist framework, the coefficients calibrated in the previous steps are either taken as fixed inputs or severely bounded. In a Bayesian fashion, the identified posterior probability density of one step becomes a prior density for the next stages [Urbina et al., 2012, Ye et al., 2022].

The sequential approach requires to define a sequence of calibration subproblems, which must be performed carefully. Indeed, as each step depends on the previous ones, an error at one step may flaw the whole calibration process. In some cases, the decomposition in subproblems can be rather easy to perform with the help of expert knowledge [Jung et al., 2022, Youn et al., 2011]. For example, Jung *et. al* [Jung et al., 2022] choose to separate the model parameters between those that describe the electrical and mechanical behaviors of an energy harvester. Yet, in some cases, such a decomposition is not available. Consequently, to avoid an arbitrary choice, another option consists in relying on a statistical criterion. For instance, given the available experiences, it can be decided to select the parameters to which the model output is sensitive following the analysis of the rank of the Fisher Information Matrix (FIM, the Hessian matrix of the log-likelihood) [Kimand and Youn, 2019]. This takes advantage of the fact that, for classical calibration problems, assuming a Gaussian discrepancy $\boldsymbol{\xi}$ in Eq. (4.1c) makes the negative log-likelihood equivalent to least squares, and in this case, the Fisher Information Matrix can be computed from the derivative of the material behavior law with respect to the model parameters. The principal limitation of these previous studies is that they do not consider the inherent variability of the individuals on which the experiments are performed (and thus they do not require a statistical model of the population). On the contrary, in this chapter, the calibration is carried out in the context of the population approaches, which complicates the modeling and the estimation procedure.

This chapter aims to adapt the sequential calibration strategy to the mixed-effects models. According to Sklar's theorem [Sklar, 1959], there exists a unique decomposition of any probability density function (PDF) between marginals and copula. Applied to the model parameters distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$, it gives :

$$f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) = c(F_1(\theta_{i1}|\boldsymbol{\Pi}_1), \dots, F_d(\theta_{id}|\boldsymbol{\Pi}_d)|\boldsymbol{\Pi}_{\text{cop}}) \times \prod_{l=1}^{d} f_l(\theta_{il}|\boldsymbol{\Pi}_l) \ , \tag{4.13}$$

with $\boldsymbol{\Pi}_{\text{cop}}$ the copula parameters, $\boldsymbol{\Pi}_l$ the parameters of the l^{th} marginal with PDF $f_l(\cdot)$ and cumulative density function $F_l(\cdot)$, θ_{il} the l^{th} coordinate of the i^{th} individual parameter $\boldsymbol{\theta}_i$, $(i,l) \in [\![1,n]\!] \times [\![1,d]\!]$ and $c(\cdot)$, the PDF of the copula (Gaussian in this chapter). In a case where all parameters are independent (the copula density is constant equal to one), following Eq. (4.13), it is possible to calibrate separately the marginals of the relevant parameters for each type of experiments. Note that in the presented procedure, the parameters whose dependencies cannot be estimated (because they do not act simultaneously on the same test) are set as independent. Furthermore, the Gaussian copula (chosen in this chapter) is compatible with a sequential scheme. Indeed, remember a Gaussian copula is parameterized by a correlation matrix which is symmetric positive-definite. Sylvester's criterion [Gilbert, 1991] states that if the determinant of the sub-matrices extracted from the full correlation matrix by deleting a row and the symmetric column is strictly positive, then the full correlation matrix is positive-definite. This condition is fulfilled if these sub-matrices are themselves positive-definite. Then, it is possible to sequentially estimate a Gaussian copula, allowing, along with the estimation of the corresponding marginals, a sequential calibration of $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$. The sequential calibration process consists in defining a sequence of subsets of indices $\mathcal{S}_k \subset [\![1,d]\!]$ corresponding to the parameters to which the model parameters is sensitive (determined for instance from the FIM) defining the marginals and correlations parameters $\boldsymbol{\Pi}_k$ to be calibrated at step k, leading to the definition of a specific likelihood function $\mathcal{L}^k$. Note that these subsets are not necessarily disjoint. In practice, it is not likely that all parameters will be sensitive enough to be properly estimated with standard experiments. Thus, it can be considered that in each step, the joint distribution of a part of the model parameters will be estimated, reducing the number of population parameters to calibrate at each step. Once these parameters are estimated, a trust region $\mathcal{R}_k$ is defined. If some of the parameters of the k^{th} step are re-estimated, they are limited to their trust region $\mathcal{R}_k$ (this choice is discussed further below). The differences between a straightforward application of the mixed-effects methods and the sequential strategy is illustrated in Figure 4.1. For the sake of clarity, l equals here to 3 (the joint distributions can be inferred from 3 types of different experiments). It is supposed that there are shared parameters between cases 1 and 2 but not with case 3 (that can be therefore estimated independently from the others). The population parameters can be divided into 3 subsets $\boldsymbol{\Pi}_1, \boldsymbol{\Pi}_2$ and $\boldsymbol{\Pi}_3$ respectively associated to data $\widetilde{\mathbf{Y}}_1$, $\widetilde{\mathbf{Y}}_2$ and $\widetilde{\mathbf{Y}}_3$. In this illustration, $\boldsymbol{\Pi}_1$ is estimated before $\boldsymbol{\Pi}_2$ (it can also be done in the reverse order). Note that there is no guarantee that the calibration in the different orders give the same results.

The sequential strategy has 2 major advantages. First, as mentioned earlier, in each step of the process, a smaller number of parameters is considered contrary to the standard counterpart, which directly decreases the difficulty of the calibration problem. Furthermore, such a strategy not only allows to ease the overall likelihood optimization as it focuses on a specific regions of the parameter space, but also enables to make sure that the calibrated distribution remains consistent with the data used in the previous stages. The definition of the sequence of calibration relies on the observation that the identification of a given marginal requires the model output to be sensitive enough to the corresponding parameters. These relevant experiments can be designed to estimate properly the PDF of interest. Algorithm 2 sums up the different stages of the proposed method.

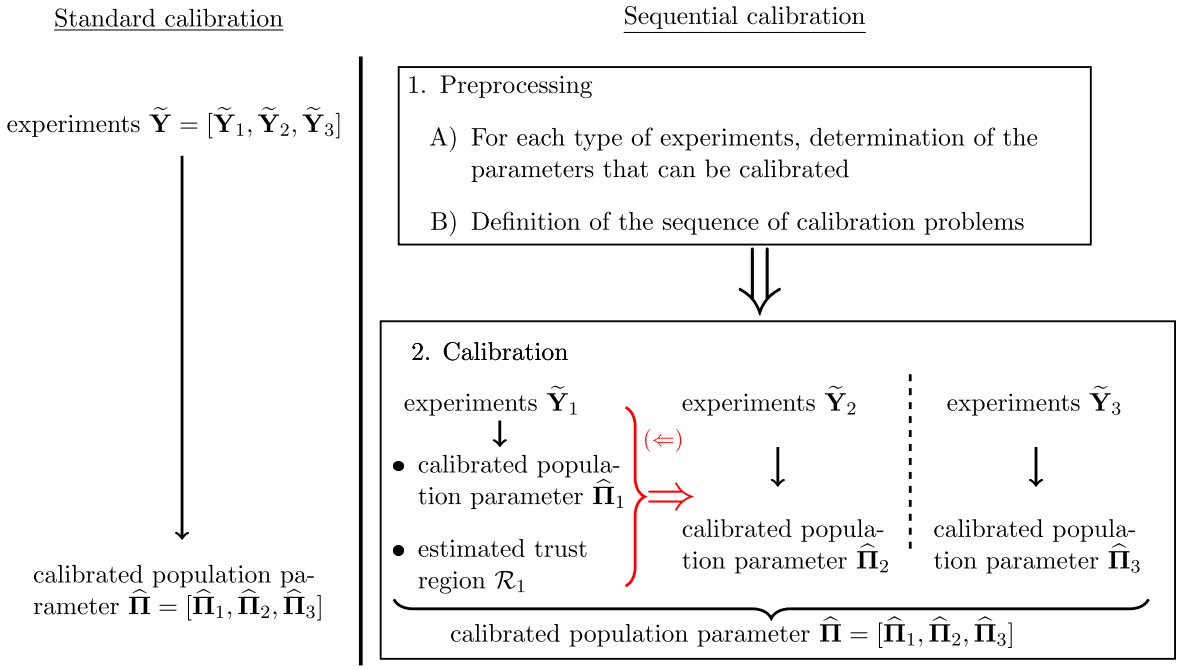


Figure 4.1: Different strategies for the calibration of the population parameters. On the left, standard mixed-effects strategy that calibrates all population parameters at once. On the right, illustration of the sequential strategy that estimates separately the population parameters associated to the considered experiments. Population parameters $\mathbf{\Pi}_3$ are estimated separately from $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$. Here, $\mathbf{\Pi}_1$ is estimated before $\mathbf{\Pi}_2$. The trust region associated to $\mathbf{\Pi}_1$ is labeled $\mathcal{R}_1$. The three calibrations performed sequentially provide an estimate of the model parameters distribution.

Algorithm 2 Algorithm of the sequential approach.$\#E$ denotes the cardinality of E.

Require: Behavior law $\mathcal{F}(\cdot,\cdot)$, available experiments $(\mathbf{Y}_i)_{i\in[\![1,n]\!]}$, α and β to define the trust regions

For all experiments, determine the model parameters that are activated on the different experiments following an analysis of the material model of interest
Group together the experiments on which are activated to the same model parameters. N_G groups are thus formed to which are attached the index of the corresponding parameters $[\mathcal{S}_k]_{k\in[\![1,N_G]\!]}$ and likelihoods $[\mathcal{L}^k(\cdot)]_{k\in[\![1,N_G]\!]}$
Maximize likelihood $\mathcal{L}^1$ within search space $\mathbb{R}^{\#\mathcal{S}_1}$ and estimate trust region $\mathcal{R}_1$
for $k = 2, \ldots, N_G$ **do**
    if any of parameters of $\mathcal{S}_k$ has already been estimated **then**
        Set search space to $\mathbb{R}^{\#\mathcal{S}_k} \cap \bigcap_{u_k \leqslant k-1} \mathcal{R}_{u_k}$, with u_k referring to the index of the stages in which some of the parameters to calibrate have already been estimated
    else
        Set search space to $\mathbb{R}^{\#\mathcal{S}_k}$
    end if
    Maximize likelihood $\mathcal{L}^k$ within determined search space
end for

The choice of the trust region to which the parameters are limited is a key feature of the process. In theory, it should illustrate the trust we have in the estimated parameters and to keep some consistency between the population parameters estimates throughout the different steps. From

a numerical point of view, the trust regions ease the negative log-likelihood minimization as the search space in which the population parameters are refined at step k should be close to the Maximum likelihood Estimate (MLE) of step $k-1$. Nevertheless, as likelihood functions may present local optima, one drawback of this method is that the search space can be reduced around a local optimum from which it can be hard to escape, making the minimization of the negative log-likelihood more prone to local convergence than in usual cases.

Several options can be explored to choose the trust region. The easiest one is an hypercube centered on the population parameters that have already been estimated. For instance, from step $k-1$ to k, it would correspond to $[\alpha\Pi_{1,k-1}, \beta\Pi_{1,k-1}] \times \cdots \times [\alpha\Pi_{s_{k-1},k-1}, \beta\Pi_{s_{k-1},k-1}]$ for the parameters $l \in [\![1, s_{k-1}]\!]$, with α and β indicating the size of the interval, s_{k-1} the number of common parameters from step $k-1$ to k and $[\Pi_{l,k-1}]_{l\in[\![1,s_{k-1}]\!]}$ the corresponding coordinates. The size of the intervals α and β can also depend on the step k and be different for each coordinate. Most of the time, α is strictly smaller than β, meaning that the $[\Pi_{l,k-1}]_{l\in[\![1,s_{k-1}]\!]}$ can be updated. Still, they can be set equal, which is tantamount to fixing the population parameters to their current estimate. In such a case, the corresponding individual parameters should not be held fixed to their mean estimate for instance but still identified (from fixed population parameters). Indeed, in that situation, the estimation of the new parameters would be dramatically dependent of the previous calibrated values. In particular, it would introduce bias in the estimation of the population parameters characterizing the dispersion of the model parameters as the number of degrees of freedom to explain the variability that arises through the samples would be in that case artificially decreased. If this trust regions management is easy to implement in practice, its main drawback is that it is arbitrary as it defines a range of variation which does not necessarily reflect the level of uncertainty of the population estimates.
A more appropriate choice for tuning this trust region would be a confidence interval, given either by the asymptotic theory or by bootstrap [Lavielle, 2014, Pinheiro and Bates, 2002]. These proposals are relevant as they quantify the uncertainty on the population parameters that comes from the lack of data in addition to other sources of errors. Thus, they seem to be the most appropriate candidates for the trust region. However, these confidence intervals techniques suffer from several limitations that make them difficult to use in practice. Indeed, aside from their respective issues discussed later, both of them are based on asymptotic assumptions, that is to say when the number of individuals increases to infinity. Of course, the asymptotic regime is not achievable because for times and costs considerations, and in practice, few individuals are available in general (in the order of 10). Thus, the intervals that reflect the uncertainty could be wide, limiting the interest of a sequential approach based on a limitation of the search space. Furthermore, both techniques suffer from their own drawbacks. For instance, the asymptotic theory is based on the FIM that is the expectation with respect to the population parameters of the second-order derivative of the log-likelihood and demands appropriate estimation techniques [Mentré et al., 1997, Vong et al., 2014]. The observed FIM, easier to estimate, can be used, at the expense of accuracy problems. This could potentially lead to numerical issues, given that the trust region is described by a covariance matrix, computed as the inverse of the FIM (that brings its own errors). Bootstrap does not suffer form this limitation as, from the bootstrap samples, confidence intervals can be estimated straightforwardly, without the need of any approximation. Yet, the accuracy of the estimation requires as much samples as possible, which is synonymous of a heavy computational burden because each sample is the result of a likelihood maximization. Finally, recall that the level of confidence that determines the size of the trust-region is user-defined, and therefore, both of these methods remain arbitrary to some extent. Consequently, for the implementation of the proposed approach, the chosen trust region is the first proposed, that is to say defined from arbitrary intervals centered on the population estimates. It remains to choose the range characterized by parameters α and β. Such settings should be considered as hyper-parameters of the method and their choice is not to be neglected.

In this work, we tried to select values that allowed to make a compromise between exploration and intensification for the maximization of the likelihood. After several trials, taking α and β respectively equal to 0.8 and 1.2 (the boundaries correspond to $\pm 20\%$ around the calibrated values) seemed to fulfill the requirements and this setting is kept in the following.

4.4 Presentation of the application case

This section presents the application of the proposed sequential calibration to the identification of an orthotropic nonlinear elastic model. After a brief description of the model in Section 4.4.1, the applied sequential strategy is exposed in Section 4.4.2 before describing the assessment protocol in Section 4.4.3.

4.4.1 A nonlinear elastic model for 2D orthotropic materials

This section aims to present the orthotropic elastic behavior of an elementary unidirectional (UD) ply in plane stress [Berthelot, 2012]. Let's note $\boldsymbol{\sigma} := (\sigma_{11}, \quad \sigma_{22}, \quad \sigma_{12})$ the Cauchy stress (in MPa) and $\boldsymbol{\varepsilon} := (\varepsilon_{11}, \quad \varepsilon_{22}, \quad 2\varepsilon_{12})$ the observed strain dimentionless within the material frame characterized by the orthogonal frame $(1, 2, 3)$. Axis 1 represents the fiber direction, axis 2 represents the matrix direction as depicted below in Figure 4.2.

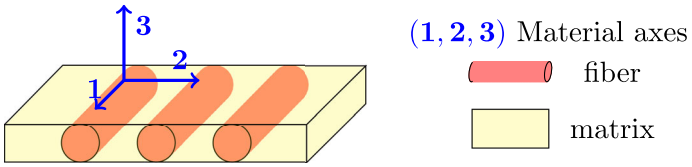


Figure 4.2: Material axes associated to the base ply

The orthotropic elastic model can be written (using the Voigt notation [Berthelot, 2012]) in the material axes :

$$\begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ 2\varepsilon_{12} \end{pmatrix} = \underline{\underline{\mathbf{S}}} \begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{12} \end{pmatrix} = \begin{pmatrix} S_{11} & S_{21} & 0 \\ S_{12} & S_{22} & 0 \\ 0 & 0 & S_{66} \end{pmatrix} \begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{12} \end{pmatrix} , \tag{4.14}$$

with $\underline{\underline{\mathbf{S}}}$ the compliance tensor. $\underline{\underline{\mathbf{S}}}$ is a symmetric tensor so $S_{12} = S_{21}$. For practical interpretations, the coefficients of the compliance tensor can be related to the elastic moduli, that is to say the longitudinal modulus E_{11}, the transverse modulus E_{22}, the shear modulus G_{12} and the Poisson's ratio ν_{12} [Berthelot, 2012].

The model presented above is elastic linear and can be used either in tension and compression. Yet, it has been observed that for some materials, in tension, 0° degree laminate tends to stiffen, while in compression, 0° degree laminate tend to soften [Laurin, 2005]. This is the consequence of the fiber behavior. Indeed, when produced, the fibers are not always properly aligned and when loaded in tension, fibers get aligned increasing the stiffness, while in compression, the initial defaults are increased and tend to soften the material. Thus, the apparent modulus increases (nonlinearly) throughout the tension experiments and conversely on the compression experiments. Thus, the longitudinal compliance S_{11} is changed and expresses as follows:

$$S_{11}^L = S_1^L + \frac{(S_{11}^0 - S_1^L)\varepsilon_0^L}{(S_{11}^0 - S_1^L)\sigma_{11} + \varepsilon_0^L} , \tag{4.15}$$

with S_{11}^0 the elastic linear compliance (module at the origin), S_1^L the asymptotic compliance and ε_0^L the abscissa of the intercept between the abscissa axis of the asymptotic curve. This interpretation comes in fact from the analysis of the expression of the longitudinal compliance. For instance, when $\sigma_{11} \to \pm\infty$, $S_{11}^L = S_1^L$ which corresponds to the asymptotic regime. On the contrary, when $\sigma_{11} = 0$, $S_{11}^L = S_{11}^0$ which corresponds to the elastic linear behavior. The symbol L indicates whether tension (T) or compression (C) is considered. The transverse compliance

S_{12}^L should also be updated and refers to $S_{12}^L \coloneqq -\nu_{12} S_{11}^L$. Figure 4.3 details the role of the different parameters.

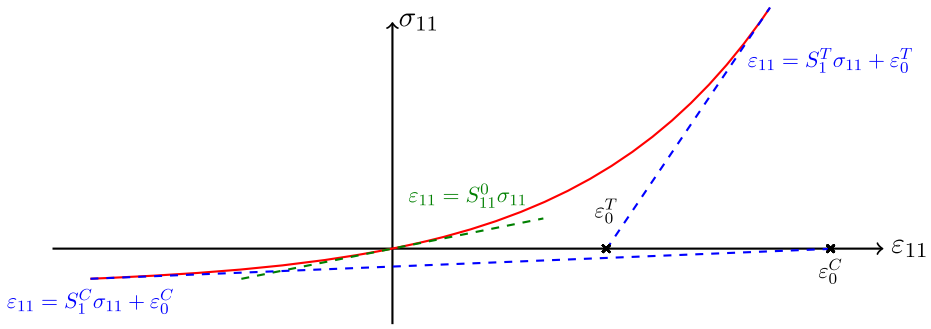


Figure 4.3: Illustration of the role of the model parameters. The solid line indicates the model output, the blue dashed lines the asymptotes and the green dashed line the tangent at the origin. Exponent T stands for tension load and C for compressive load.

As the compliance of the material decreases in tension and increases in compression, the parameters should comply with $S_1^T < S_{11}^0$ and $S_1^C > S_{11}^0$. With this parameterization, the model is not identifiable. In fact, it is over-parameterized, *i.e.* there are several combinations of both S_1^L and ε_0^L that give the same output as shown by Germain [Germain, 2020]. Thus, ε_0^L is fixed to a usual nominal value, here 0.005 [Germain, 2020]. Finally, there are 6 parameters to be calibrated $S_1^T, S_1^C, S_{11}^0, S_{22}, \nu_{12}$ and S_{66}.
In theory, the compliance matrix $\underline{\underline{\mathbf{S}}}$ should be positive-definite [Berthelot, 2012]. To enforce this property on the individual compliance tensor, two requirements have to be met: the strict positivity of S_{66} and the positive-definiteness of the submatrix containing all other compliances (this is in fact the application of the Sylvester criterion [Gilbert, 1991]). The first point is easy to check but the second one deserves more attention. Indeed, it can only be guaranteed if all the involved compliances can be calibrated on the same specimen. In that case, it is possible to enforce such property by implementing constraints between the compliances or by decomposing the compliance matrix with a relevant factorization (*e.g.,* Cholesky, spherical decomposition). However, none of these solutions is fully satisfactory as they either increase the difficulty of the auxiliary minimizations or use parameters that are difficult to interpret from a physical point of view. In any other case, for most calibration methods, it is not possible to guarantee positive-definitness of the compliance matrix in a simple way, because, on a given specimen, all the parameters of interest are not available. A more convenient way to proceed, chosen in this paper, consists in calibrating the elastic behavior without accounting for this constraint and to discard the realizations of matrices that are not definite positive in a subsequent step.

4.4.2 Definition of the sequential strategy

Given the modeling choices, the distribution of the model parameters is described by 27 parameters, namely 6 mean parameters, 6 variance parameters and 15 covariance parameters. The distribution parameters can be summed up into a mean vector $\boldsymbol{\mu} = \begin{pmatrix} \mu_{S_{11}^0} & \mu_{S_1^T} & \mu_{S_1^C} & \mu_{\nu_{12}} & \mu_{S_{22}} & \mu_{S_{66}} \end{pmatrix}^\top$ and covariance matrix:

$$\boldsymbol{\Sigma} = \begin{array}{c} \\ S_{11}^0 \\ S_1^T \\ S_1^C \\ \nu_{12} \\ S_{22} \\ S_{66} \end{array} \begin{array}{c} \begin{array}{cccccc} S_{11}^0 & S_1^T & S_1^C & \nu_{12} & S_{22} & S_{66} \end{array} \\ \begin{pmatrix} \mathbb{V}(S_{11}^0) & * & * & * & * & * \\ \mathrm{Cov}(S_1^T, S_{11}^0) & \mathbb{V}(S_1^T) & * & * & * & * \\ \mathrm{Cov}(S_1^C, S_{11}^0) & \mathrm{Cov}(S_1^C, S_1^T) & \mathbb{V}(S_1^C) & * & * & * \\ \mathrm{Cov}(\nu_{12}, S_{11}^0) & \mathrm{Cov}(\nu_{12}, S_1^T) & \mathrm{Cov}(\nu_{12}, S_1^C) & \mathbb{V}(\nu_{12}) & * & * \\ \mathrm{Cov}(S_{22}, S_{11}^0) & \mathrm{Cov}(S_{22}, S_1^T) & \mathrm{Cov}(S_{22}, S_1^C) & \mathrm{Cov}(S_{22}, \nu_{12}) & \mathbb{V}(S_{22}) & * \\ \mathrm{Cov}(S_{66}, S_{11}^0) & \mathrm{Cov}(S_{66}, S_1^T) & \mathrm{Cov}(S_{66}, S_1^C) & \mathrm{Cov}(S_{66}, \nu_{12}) & \mathrm{Cov}(S_{66}, S_{22}) & \mathbb{V}(S_{66}) \end{pmatrix} \end{array} , \tag{4.16}$$

where the $*$ stands for the symmetric coefficient of the covariance matrix.

To identify the elastic constants, a usual practice is to rely on tests on unidirectional $0°, 90°$ and $\pm 45°$ laminates [Berthelot, 2012, Daniel and Ishai, 2006, Tomblin et al., 2001]. Indeed, from a physical point of view, the behavior of $0°$ laminate can be exclusively described by S_{11}^0, S_1^T, S_1^C that characterize the fiber behavior and ν_{12} which also takes into account the phenomena that occur within the matrix. Similarly, the linear part of the longitudinal response of a tension/compression test on a $90°$ laminate is mainly controlled by S_{22}, the same applies for the linear part of the difference between the longitudinal and transverse responses for the $\pm 45°$ laminate, mostly driven by S_{66} [Berthelot, 2012, Daniel and Ishai, 2006, Tomblin et al., 2001]. Remember that, with mixed-effects, to estimate properly a given marginal, the corresponding parameters should be accurately estimated over the different identified individuals, and the same occurs for the parameters involved in statistical dependency relation. Thus, from the analyses made earlier, the correlations between $S_{11}^0, S_1^T, S_1^C, \nu_{12}$ and S_{22} cannot be estimated with the usual elementary experiments, and the same applies to those between $S_{11}^0, S_1^T, S_1^C, \nu_{12}$ and S_{66} or between S_{22} and S_{66}. In fact, from these elementary experiments, only the joint distribution of $S_{11}^0, S_1^T, S_1^C, \nu_{12}$ and the marginals of S_{66} and S_{22} can be estimated, which gives three distributions to calibrate (in addition to the residual models). This makes 18 population parameters (14 for the joint distribution of $S_{11}^0, S_1^T, S_1^C, \nu_{12}$, and 2 for both marginals of S_{22} and S_{66}) to estimate.

From this assessment, two solutions arise. On the one hand, it is possible to calibrate all the experiments together by making sure that for each individual, only the relevant individual parameters are calibrated. On the other hand, it is possible to apply the sequential strategy dividing the estimation of the 18 population parameters into three subproblems. These two solutions are detailed below, that correspond respectively to the left and right situations depicted in Figure 4.1.

To estimate the 18 population parameters by resolving a single calibration problem, design matrices introduced in Eq. (4.3) $\mathbf{A}_i$ and $\mathbf{B}_i$ can be employed. For instance, with the mean vector defined earlier, one possible choice would be $\mathbf{A}_i = \mathbf{B}_i$ equal to the identity matrix of dimension 6 to which the last two rows would be deleted (and $\mathbf{A}_i$ would have thus 4 rows with 6 columns) for the UD $0°$ experiments, $\mathbf{A}_i = \mathbf{B}_i = (0, 0, 0, 0, 1, 0)^\top$ for the UD $90°$ experiments and $\mathbf{A}_i = \mathbf{B}_i = (0, 0, 0, 0, 0, 1)^\top$ for the $\pm 45°$ experiments. This solution refers to the situation depicted in the left part of Figure 4.1.

However, here the calibration process can be significantly eased, as the parameters that describe the behavior of the $0°$ laminates do not express themselves on the behavior of the $90°$ laminates

and similarly between the other laminates. It is therefore possible to identify the joint distribution of $S_{11}^0, S_1^T, S_1^C, \nu_{12}$ and the two other marginals independently (which thus do not require to use design matrix). If the estimation of these distributions is carried out separately, this can be considered as a first application of the sequential strategy, given that between the three cases, there are not shared parameters, which can therefore be studied independently. This solution refers to the situation depicted in the right part of Figure 4.1. The calibration of the marginals of S_{66} and S_{22} are carried out from experiments with univariate observations that exhibit a linear behavior. These cases only refer to univariate linear mixed-effects models extensively discussed in [Demidenko, 2013, Lavielle, 2014, Pinheiro and Bates, 2002]. Consequently, the study is here focused on the calibration of the joint distribution of S_{11}^0, S_1^T, S_1^C and ν_{12}.

One first remark that can be made is that S_1^C is active only when compressive loads are applied to the sample, and S_1^T is active only when tension loads are applied. Thus, from the elementary tests, there is no configuration in which they are activated simultaneously, and thus the correlation between them cannot be estimated. Therefore, contrary to the above comments, the joint distribution of S_{11}^0, S_1^T, S_1^C and ν_{12} is parameterized by 13 and not 14 population parameters. S_{11}^0 and ν_{12} are activated both in tension and compression. Note that ν_{12} is expressed only through the transverse responses, either in tension and compression. There are shared parameters for the two types of experiments, S_{11}^0 and ν_{12}. Design matrices can be used. For instance, for the tension experiments, matrices $\mathbf{A}_i$ and $\mathbf{B}_i$ would correspond to the identity matrix of dimension 4 to which is deleted the third row associated to S_1^C (so both have 3 rows and 4 columns) and for the compression experiments, they could be taken as the identity matrix of dimension 4 to which is deleted the second row associated to S_1^T (so both have 3 rows and 4 columns).

The interest of this test case is that it can be used to illustrate the sequential strategy. In particular, it enables to compare the accuracy of the calibration from the sequential strategy to the solution that implements design matrices. For 0° laminates, let's consider a tension load $\boldsymbol{\sigma} = \begin{pmatrix}\sigma_{11}, & 0, 0\end{pmatrix}$ with $\sigma_{11} > 0$. Then, following Eq. (4.14), the model output is determined by S_{11}^0, S_1^T and ν_{12}. It is thus possible to calibrate the parameters highlighted in red in Eq. (4.17):

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_{S_{11}^0} \\ \mu_{S_1^C} \\ \mu_{S_1^T} \\ \mu_{\nu_{12}} \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{array}{c} \\ S_{11}^0 \\ S_1^C \\ S_1^T \\ \nu_{12} \end{array} \begin{array}{c} \begin{array}{cccc} S_{11}^0 & S_1^C & S_1^T & \nu_{12} \end{array} \\ \begin{pmatrix} \mathbb{V}(S_{11}^0) & * & * & * \\ \mathrm{Cov}(S_1^C, S_{11}^0) & \mathbb{V}(S_1^C) & * & * \\ \mathrm{Cov}(S_1^T, S_{11}^0) & 0 & \mathbb{V}(S_1^T) & * \\ \mathrm{Cov}(\nu_{12}, S_{11}^0) & \mathrm{Cov}(\nu_{12}, S_1^C) & \mathrm{Cov}(\nu_{12}, S_1^T) & \mathbb{V}(\nu_{12}) \end{pmatrix} \end{array}. \tag{4.17}$$

It is possible here to estimate a first trust region : $\mathcal{R}_1$. Let's now consider a compressive load $\boldsymbol{\sigma} = \begin{pmatrix}\sigma_{11}, & 0, & 0\end{pmatrix}$ with $\sigma_{11} < 0$. Then, following Eq. (4.14), the model output is determined by S_{11}^0, S_1^C and ν_{12}. It is possible to calibrate the parameters highlighted in blue in Eq. (4.18):

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_{S_{11}^0} \\ \mu_{S_1^C} \\ \mu_{S_1^T} \\ \mu_{\nu_{12}} \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{array}{c} \\ S_{11}^0 \\ S_1^C \\ S_1^T \\ \nu_{12} \end{array} \begin{array}{c} \begin{array}{cccc} S_{11}^0 & S_1^C & S_1^T & \nu_{12} \end{array} \\ \begin{pmatrix} \mathbb{V}(S_{11}^0) & * & * & * \\ \mathrm{Cov}(S_1^C, S_{11}^0) & \mathbb{V}(S_1^C) & * & * \\ \mathrm{Cov}(S_1^T, S_{11}^0) & 0 & \mathbb{V}(S_1^T) & * \\ \mathrm{Cov}(\nu_{12}, S_{11}^0) & \mathrm{Cov}(\nu_{12}, S_1^C) & \mathrm{Cov}(\nu_{12}, S_1^T) & \mathbb{V}(\nu_{12}) \end{pmatrix} \end{array}. \tag{4.18}$$

In the second step, $\mu_{S_{11}^0}, \mu_{\nu_{12}}, \mathbb{V}(S_{11}^0), Cov(S_{11}^0, \nu_{12})$ and $\mathbb{V}(\nu_{12})$, which correspond to parameters active in both experiments, are limited to their trust region $\mathcal{R}_1$. All the distribution parameters of the model are now estimated through these two steps.

From this point, it could be decided to restart the calibration with the tension experiments to refine the estimation of the joint distribution of S_{11}^0, S_1^T and ν_{12} by limiting the parameters of the joint distribution of S_{11}^0, ν_{12} to trust region $\mathcal{R}_2$. This would be repeated upon reaching convergence, as in a fixed point problem. The first difficulty is to define convergence (at the level of the population parameters, of the calibrated distributions, *etc.*). This is of prime importance because a stationary behavior could only be met for some convergence criteria, and nothing guarantees that only few iterations would be enough, increasing the computational costs. Still, this chapter is focused on the sequential strategy, leaving the description and the generalization of the sequential strategy with a fixed-point approach to future investigations.
The application of the sequential strategy has been presented in one way, starting with the tension experiments and ending with the compression experiments. Yet, there is no particular justification to such a choice. Consequently, it is here necessary to carry out the sequential strategy in both ways, starting with the tension experiments and ending with the compression experiments and then starting with the compression experiments and ending with the tension experiments. In the rest of the chapter, the results associated to the calibration starting with the tension experiments and ending with the compression experiments are presented in green, those coming from the calibration starting with the compression experiments and ending with the tension experiments are presented in black and those considering both types of experiments in red. It is possible to design complex experiments (for instance [Gogu, 2009]) from which the contribution of each parameter can be extracted. Still, because of the complexity of the tests, the calibration is significantly harder to perform, especially because it can become necessary to treat field data. Furthermore, in practice, as they are harder to make, fewer individuals are available compared to the elementary tests.

Remember that laminates are composed of several plies. The approach developed above assumed that all plies were described by the same individual elastic properties. This is a strong assumption but appears necessary to ensure the appliability of the method. Indeed, if all plies have the same properties, the 0° laminate can be considered as composed of only one ply and the individuals are characterized by 3 parameters, namely $S_{11}^0, S_{11}^L, \nu_{12}$. However, if we consider that for each individual, the properties of all plies are different, for a 0° laminate with 8 plies, each specimen would be characterized by $8 \times 3 = 24$ parameters. In total, this makes n optimization problems of function $g_i(\cdot)$ defined in Eq. (4.9) in dimension 24 which may not guarantee a proper estimation of the individual parameters and thus flaw the estimation of the population parameters. If the plies used to build the laminates are taken within the same sheet, this assumption can be considered as acceptable given the gradient of the material properties within a composite sheet is usually negligible. However, if the plies used to build the final laminate are taken from different sheets, this hypothesis can be challenged as the plies can be considered as different realizations of the material properties. In such a case, this approach may reduce the variability of the elastic constants. The consequences of such a choice will be studied further in the propagation phase carried out in Section 4.7.

One of the main advantages of the sequential approach is to reduce the maximum number of parameters to estimate compared to the standard method using design matrices. For instance, the calibration of the joint distribution of the compliances for the nonlinear elasticity, parameterized by 17 parameters (6 mean parameters, 6 standard deviations and 5 correlations as detailed above) with a dependence structure and 12 without (6 mean parameters and standard deviations), is divided with a sequential scheme into subproblems with at most 9 parameters with a dependence structure and 6 without. This reduction can be illustrated further if another component is added, for instance the viscoelasticity behavior as in the OPFM model [Laurin, 2005]. This component adds 6 new parameters and can be calibrated from both 90° and ±45° laminates (in which also express the linear compliances S_{22} and S_{66}). With the same modeling

assumptions (Gaussian marginals and copula), the joint distribution is parameterized with 42 parameters with a dependence structure (all correlations but two can be estimated) and 16 without. Implementing a sequential scheme allows to substitute to these problems two calibration of distributions with respectively 27 parameters with a dependence structure and 12 without. This analysis does not take into account the residual models which is discussed later. Table 4.1 recaps the number of population parameters to estimate with the sequential approach or using design matrices for both elasticity and viscoelasticity behaviors considering a dependence structure or not.

Table 4.1: For the nonlinear elastic and viscoelastic behavior laws of the OPFM model, number of population parameters to be calibrated with design matrices or using the sequential approach considering a dependence structure or not. The residual models are not taken into account. For the sequential approach, the number of population parameters refers to the highest number of parameters to be estimated in the different stages of the method.

Calibration method	Number of population parameters for the 2D nonlinear elasticity (see Eq. (4.4.1))	Number of population parameters for the 2D viscoelasticity behavior
Design matrices with dependence structure	17	42
Design matrices without dependence structure	12	16
Sequential approach with dependence structure	9	27
Sequential approach without dependence structure	6	12

A first remark that can be made from Table 4.1 is the interest to separate the identification of the parameters between the different phenomena. Furthermore, it shows that the sequential approach allows to limit the number of population parameters to calibrate for each phenomena at each iteration. This is particularly true for the viscoelasticity phase, in which the implementation of the sequential approach with a dependence structure cuts almost by half the maximum number of population parameters that are estimated in the same problem, even if this remains difficult to carry out. However, it should be emphasized that standard mixed-effects do not enable to calibrate more population parameters than the sequential strategy. At the end of the calibration process, both methods provide the same distributions characterized by the same number of population parameters, statistical dependencies included. Table 4.1 also illustrates that a dependence modeling inflates the number of population parameters. This is expected because with statistical dependencies, there are more population parameters to estimate. Thus, because of the small number of available individuals, striking a balance between the complexity of the statistical modeling and the number of available individuals appears as necessary, provided that the correlation coefficients demand a large amount of data to be accurately estimated. So far, only the number of distribution parameters to estimate has been discussed. Yet, with mixed-effects, the residual model (here the ω_is) is also to calibrate. Here again the implementation of the sequential approach reduces the number of residual models (and thus of parameters) to estimate. Indeed, with the sequential approach, a focus is made on certain types of specimen (for instance compression experiments) while with design matrices, different types of experiments are considered (tension and compression experiments). Thus, under the reasonable

hypothesis of at least one residual model for each type of experiments, there are more residual models (so more parameters) to calibrate when design matrices are employed than in the different stages of the implementation of the sequential approach. Consequently, it appears that the sequential strategy is particularly suited to reduce the complexity of the minimization problem. The interest of this test case is that it the sequential strategy can be tested and compared to its direct counterpart. The next section describes the tools used to analyze the results. It should be noticed that, in this application of the sequential strategy, the calibration subproblems have been defined from the analysis of the available experiments, allowing to select the parameters of model that would be used to describe the phenomena of interest.

4.4.3 Testing protocol for mixed-effects calibration

The testing protocol of mixed-effects is presented here. The generation of the virtual data is detailed first, then the numerical settings for the optimization of the opposite log-likelihood, to end with a recap of the testing protocol.

4.4.3.1 Generating virtual data

A multivariate distribution, defined by population parameters $\mathbf{\Pi}$, is chosen to generate synthetic data. This set of parameters will be denoted $\mathbf{\Pi}_{\text{exact}}$ in the following. Individual parameters $\boldsymbol{\theta}_{i,\text{exact}}$ are sampled from this distribution to compute, for each of these specimens, the nonlinear elastic strain outputs $\boldsymbol{\varepsilon}_{i,\text{exact}}$. A noise profile is added to this data to obtain the synthetic data involved in the calibration process $\boldsymbol{\varepsilon}_{i,\text{noisy}}$. The mixed-effects approach is then applied to the synthetic data and the results obtained are confronted to their exact counterpart.
As in Chapter 3, this model parameters distribution has been chosen as a multivariate Gaussian distribution, which corresponds to a usual modeling [Bright et al., 2011b, Niyigena et al., 2016] and such a modeling is kept in this chapter. Such a distribution is parameterized by both a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\mathbf{\Sigma}$. Table 4.2 presents the values of the latter to define the multivariate Gaussian distribution used to generate the synthetic data. Mean values are consistent with material properties of composite material with carbon fibers and epoxy resin such as those described in Laurin [Laurin, 2005]. Standard-deviations of S_{11}^0 and ν_{12} are chosen to fit with the values of dispersion given by Rollet [Rollet, 2007]. The exact means and standard-deviations of the parameters for the joint distribution $f_{\mathbf{\Theta}|\mathbf{\Pi}_{\text{exact}}}$ are described in Table 4.2.

Table 4.2: Exact values of the parameter distribution.

	S_{11}^0 [$\times 10^{-7}$MPa^{-1}]	S_1^C [$\times 10^{-7}$MPa^{-1}]	S_1^T [$\times 10^{-7}$MPa^{-1}]	ν_{12} [$\times 10^{-2}$]
Mean	85.2	155	64.7	33.1
Standard-deviation	1.31	17.3	4.19	0.660

Several correlation scenarios are considered: independence between S_1^C and S_1^T, positive correlation : $\rho(S_{11}^0, S_1^T) = 0.809$, $\rho(\nu_{12}, S_1^C) = 0.730$ and negative correlations $\rho(S_1^T, \nu_{12}) = -0.500$, $\rho(S_{11}^0, \nu_{12}) = -0.741$, $\rho(S_1^C, S_{11}^0) = -0.414$, noting ρ the correlation. Note that these values are arbitrary and do not refer to physical values, as they are not available in the literature for the material of interest to the best of our knowledge. For each of the 20 calibration repetitions (to evaluate the robustness to the selected samples), 50 independent identically distributed vectors of parameters corresponding to 50 virtual individuals are sampled from $f_{\mathbf{\Theta}|\mathbf{\Pi}_{\text{exact}}}$ to achieve statistical consistency of the estimators. Note that 50 repetitions in an experimental context can be considered as large, but with synthetic data, it eases the analysis of the calibration results. For each of these samples, given a tension stress that goes from 0 to 1,500 MPa with 32 points regularly spaced and a compressive load that goes from 0 to −1,000 MPa with 28 points regularly spaced, the corresponding model outputs $(\boldsymbol{\varepsilon}_{i,exact})_{i\in[\![1,n]\!]}$ are computed. An heteroscedastic

noise is added to the experimental data:

$$\boldsymbol{\varepsilon}^{j}_{i,noisy} = \boldsymbol{\varepsilon}^{j}_{i,exact} \times (\mathbf{1}_{\mathbb{R}^2} + \tau\boldsymbol{\xi}) \text{ with } \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}_{\mathbb{R}^2}, \mathbf{I}_2) \text{ and } \tau = 0.02 \ , \tag{4.19}$$

with $\mathbf{1}_{\mathbb{R}^2}$ the unitary vector of $\mathbb{R}^2$, $\mathbf{0}_{\mathbb{R}^2}$ the zero vector of $\mathbb{R}^2$ and $\mathbf{I}_2$ the identity matrix of $\mathcal{M}_2(\mathbb{R})$. In the calibration process, the noise is considered as homoscedastic for all measures (*cf.* Eq. (4.20)) as otherwise, this would make too much residual models to estimate (up to 2,800 in compression and 3,200 in tension). In addition, as illustrated in the previous chapter, the homoscedastic assumption was sufficient to provide relevant estimations of the population parameters, even if at any time the heteroscedastic hypothesis can be selected. Formally, this can be written as:

$$\boldsymbol{\Omega}_i = \boldsymbol{\Omega}, \ \forall i \in [\![1, n]\!] \ . \tag{4.20}$$

The full parameters to estimate $\boldsymbol{\Psi}$ refer to $\boldsymbol{\Psi} = [\boldsymbol{\Pi}, \boldsymbol{\Omega}]$. Exact and noisy data are depicted in Figure 4.4.

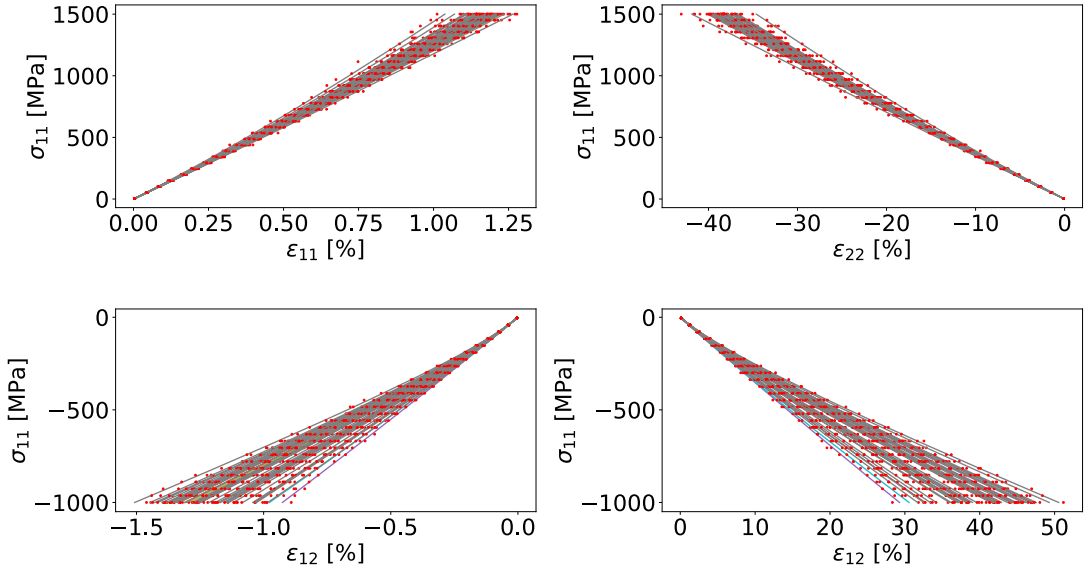


Figure 4.4: Example of a set of 50 synthetic stress-strain curves, without and with added noise (lines and dots, respectively). The longitudinal (on the left) and transverse responses (on the right) are depicted for both the tension (upper line) and compression experiments (lower line). 10 experiments out of are 50 highlighted in red for the sake of clarity.

4.4.3.2 Minimization of the opposite log-likelihood

To estimate the population parameters and the residual model all assembled within $\boldsymbol{\Psi}$, it is necessary to maximize the likelihood of mixed-effects models or rather to minimize the opposite of the log-likelihood, which can be translated into the following optimization problem:

$$\widehat{\boldsymbol{\Psi}} = \underset{\boldsymbol{\Psi} \in \Gamma_{\boldsymbol{\Psi}}}{\arg\min} - \ln(\mathcal{L}(\boldsymbol{\Psi})) \ , \tag{4.21}$$

with $\widehat{\boldsymbol{\Psi}}$ the maximum likelihood estimate of $\boldsymbol{\Psi}$ which decomposes between the population estimates $\widehat{\boldsymbol{\Pi}}$ and the calibrated residual model $\widehat{\boldsymbol{\Omega}}$ and $\Gamma_{\boldsymbol{\Psi}}$, the search space for $\boldsymbol{\Psi}$. The model parameters distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ is assumed to be multivariate Gaussian distribution (which turns out to be the exact hypothesis). Thus, the population parameters refer to means, standard-deviations and correlations. The bounds of the search space are described in Table 4.3. If these search bounds are hit, they can be updated if deemed necessary. The search bounds on the population parameters are set up so that the mid point correspond approximately to the exact

value of the population parameter. The search bounds also intend to allow an exploration of the parameter space. Those on $\boldsymbol{\Omega}$ are determined by trial-error.

Table 4.3: Search space for the population parameters.

$\mu_{S_{11}^0}$ $[10^{-7}\text{MPa}^{-1}]$	$\mu_{S_1^T}$ $[10^{-7}\text{MPa}^{-1}]$	$\mu_{S_1^C}$ $[10^{-7}\text{MPa}^{-1}]$	$\mu_{\nu_{12}}$ $[10^{-2}]$
$[10.0, 110.0]$	$[50.0, 130]$	$[60.0, 250]$	$[15.0, 50.0]$

$\text{sd}(S_{11}^0)$ $[10^{-7}\text{MPa}^{-1}]$	$\text{sd}(S_1^T)$ $[10^{-7}\text{MPa}^{-1}]$	$\text{sd}(S_1^C)$ $[10^{-7}\text{MPa}^{-1}]$	$\text{sd}(\nu_{12})$ $[10^{-2}]$
$[0.500, 4.50]$	$[0.500, 10.0]$	$[5.00, 30.0]$	$[0.100, 3.00]$

ω_{long}^T $[\times 10^{-6}]$	ω_{trans}^T $[\times 10^{-5}]$	ω_{long}^C $[\times 10^{-5}]$	ω_{trans}^C $[\times 10^{-5}]$
$[1.00, 10.0]$	$[3.16, 31.6]$	$[1.00, 10.0]$	$[3.16, 31.6]$

The same difficulties for the minimization of the opposite log-likelihood arise in this test case as those underlined in Chapter 3. The same evolutionary algorithm called the Covariance Matrix Adaptation Evolution Strategy [Hansen et al., 2003] is chosen to minimize the objective function. The optimization variables $\boldsymbol{\Psi}$ are normalized between 0 and 1. At most 400 iterations of the algorithm are carried out to maximize the likelihood. In [Hansen et al., 2003], the size of the population of the CMA-ES algorithm suggested is $\lceil 4 + 3 \times \ln(p) \rceil$ points, p being the number of full model parameters to calibrate ($\lceil \cdot \rceil$ the ceiling function). The starting point corresponds to the center of the search bounds to which is added a slight deviation, which gives as starting point $0.65 \times \mathbf{1}_{\mathbb{R}^p}$. Default settings for the convergence of the minimization are considered: 10^{-6} for the tolerance on both the objective function and the normalized $\boldsymbol{\Psi}$.

4.4.3.3 Estimation of the individual parameters

Mixed-effects calibration for phenomenological model aims to tune the population parameters to make the model parameters distribution fit the best with the empirical distribution of the individual parameters. Consequently, the chosen method should be able to identify properly the individual parameters on the different samples. For the Laplace approximation, this is performed through the minimization of function $g_i(\cdot)$ defined in Eq. (4.9). During the optimization of the global likelihood, it is difficult to assess the quality of the identification of the individual parameters. Chapter 3 demonstrated that the resolution of this minimization problem could be solved with the SLSQP algorithm from NLopt [Johnson, 2007] and implemented in OpenTURNS [Baudin et al., 2016] instead of the CMA-ES algorithm without major losses of accuracy. However, with the nonlinear elastic model, numerical issues provoking the failure of the auxiliary minimization (and of the likelihood estimation) occurred during the minimization of the opposite log-likelihood. To ensure a reliable estimation of the population parameters, the Nelder-Mead algorithm is selected. One of its main advantage is that it is gradient-free, allowing to process objective functions with limited smoothness properties (in particular it does not require it to be differentiable). Yet, as the SLSQP algorithm, it is not a global optimizer, meaning that it is sensitive to the initialization point. This aspect is solved by using the same tool, by considering the minimization in a multi-start fashion. 50 initialization points are used, whose 49 come from a LHS design from OpenTURNS [Baudin et al., 2016] without random shuffle, the last one corresponding to the mean vector. Default settings of the Nelder-Mead algorithm for the convergence of the minimization are considered. The maximum number of model evaluations is set to 1,000. The bounds are centered on the fixed-effects that refer to the most probable value for the individual parameters and authorize to explore a large number of possibilities for the individual parameters. The search space is limited around the fixed-effects that stand for the initial point. It is defined as $\Gamma_{\boldsymbol{\theta}} = [0.3\boldsymbol{\beta}, 1.7\boldsymbol{\beta}]$, with $\boldsymbol{\beta}$ the fixed-effects (which reduce to the mean for a Gaussian multivariate distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$). The optimization variables are normalized to ease the convergence.

4.4.3.4 Testing protocol

To test the ability of the proposed methodology to calibrate the elastic model with mixed-effects, it is chosen to calibrate the model with synthetic data. Both the distribution parameters and the individual parameters are known to generate the synthetic data and are estimated with the proposed calibration approach.
A multivariate Gaussian distribution, defined by a mean vector $\boldsymbol{\mu}_{\text{exact}}$ and a covariance matrix $\boldsymbol{\Sigma}_{\text{exact}}$ is chosen to generate the synthetic data. This set of parameters will be denoted $\boldsymbol{\Pi}_{\text{exact}}$ in the following. Individual parameters $\boldsymbol{\theta}_{i,\text{exact}}$ are sampled from this distribution to compute, for each of these individuals the elastic model outputs $\boldsymbol{\varepsilon}_{i,\text{exact}}$. The mixed-effects approach is applied to the synthetic data and the results obtained are confronted to their exact counterpart. The results are first compared at the population level to ensure that the model parameters distribution is properly estimated, using the relative error on the distribution parameters $\mathcal{E}(\widehat{\boldsymbol{\Pi}})$, defined as:

$$\mathcal{E}(\widehat{\boldsymbol{\Pi}}) \coloneqq \left[\mathcal{E}(\widehat{\boldsymbol{\Pi}}_1), \quad \ldots, \quad \mathcal{E}(\widehat{\boldsymbol{\Pi}}_s)\right] \text{ with } \mathcal{E}(\widehat{\boldsymbol{\Pi}}_l) \coloneqq 100 \times \frac{|\Pi_{\text{exact},l} - \widehat{\Pi}_l|}{\Pi_{\text{exact},l}}, \forall l \in [\![1, s]\!] \ , \tag{4.22}$$

noting s the number of distribution parameters and the KL divergence [Kullback and Leibler, 1951], $\text{KL}(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}_{\text{exact}}}, f_{\boldsymbol{\Theta}|\widehat{\boldsymbol{\Pi}}})$, between the exact $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}_{\text{exact}}}$ and the estimated distribution $f_{\boldsymbol{\Theta}|\widehat{\boldsymbol{\Pi}}}$:

$$\text{KL}(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}_{\text{exact}}}, f_{\boldsymbol{\Theta}|\widehat{\boldsymbol{\Pi}}}) \coloneqq \int_{\mathbb{R}^d} f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}_{\text{exact}}}(\boldsymbol{\theta}) \ln\left(\frac{f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}_{\text{exact}}}(\boldsymbol{\theta})}{f_{\boldsymbol{\Theta}|\widehat{\boldsymbol{\Pi}}}(\boldsymbol{\theta})}\right) \mathrm{d}\boldsymbol{\theta} \ . \tag{4.23}$$

In general, it is computed with coarse Monte-Carlo. Yet, if both the exact and calibrated model parameters distribution are Gaussian multivariate distribution of dimension d parameterized by $\boldsymbol{\Pi}_{\text{exact}} = [\boldsymbol{\mu}_{\text{exact}}, \boldsymbol{\Sigma}_{\text{exact}}]$ and $\widehat{\boldsymbol{\Pi}} = \left[\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}\right]$ respectively, the Kullback-Leibler (KL) divergence can be estimated in closed form and refers to [Pardo, 2018]

$$\text{KL}(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}_{\text{exact}}}, f_{\boldsymbol{\Theta}|\widehat{\boldsymbol{\Pi}}}) = \tfrac{1}{2}\left(\text{tr}(\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}_{\text{exact}}) + (\boldsymbol{\mu}_{\text{exact}} - \widehat{\boldsymbol{\mu}})^\top \widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\mu}_{\text{exact}} - \widehat{\boldsymbol{\mu}}) + \ln\left(\tfrac{|\widehat{\boldsymbol{\Sigma}}|}{|\boldsymbol{\Sigma}_{\text{exact}}|}\right) - d\right) \ . \tag{4.24}$$

The KL divergence expresses a notion of discrepancy between two distributions. One interest of such a measure is that it accounts simultaneously for all aspects (marginals and correlations) when the error on the population parameters only studies the calibration of one population parameter at the time. In addition, it is here specifically relevant as the population parameters are estimated jointly and the KL divergence can be understood as a joint measure of the errors. The adequation between the calibrated and the synthetic data can also be evaluated for each individual. To verify that the different individuals are correctly identified either in terms of calibrated individual parameters $\widehat{\boldsymbol{\theta}}_i$ or estimated output $\mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)$, the averaged relative error on the individual parameters $\mathbf{e}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_n)$ in %:

$$\begin{aligned} &\mathbf{e}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_n) \coloneqq \frac{1}{n}\sum_{i=1}^{n} \mathbf{e}_i(\widehat{\boldsymbol{\theta}}_i) \text{ with } \mathbf{e}_i(\widehat{\boldsymbol{\theta}}_i) = \left[e_{i1}(\widehat{\theta}_{i1}), \quad \ldots, \quad e_{id}(\widehat{\theta}_{id})\right] \\ &e_{il}(\widehat{\theta}_{il}) \coloneqq 100 \times \frac{|\theta_{il,\text{exact}} - \widehat{\theta}_{il}|}{\theta_{il,\text{exact}}} \ \forall l \in [\![1, d]\!] \ , \end{aligned} \tag{4.25}$$

and the average error $\mathbf{D}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_n)$ between the exact and estimated strains for each coordinate

$$\begin{aligned} &\mathbf{D}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_n) \coloneqq \left[D_1(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_n), \quad \ldots, \quad D_m(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_n)\right] \\ &\text{with } D_q(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_n) \coloneqq \frac{1}{n}\sum_{i=1}^{n} \frac{1}{N_i} \|\varepsilon^q_{i,exact} - \mathcal{F}^q(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)\|_2^2 \ \forall q \in [\![1, m]\!] \ , \end{aligned} \tag{4.26}$$

are computed. Furthermore, to account for the specificity of the sampled individuals, the calibration is repeated 20 times with different individuals generated from the same joint distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}_{\text{exact}}}$. The above indicators are averaged over the repetitions and their dispersion is characterized by the coefficient of variation $\text{COV}(X)$ defined for any non-zero quantity X as $\text{COV}(X) = \frac{\sqrt{\widehat{V}(X)}}{\widehat{M}(X)}$ with $\widehat{M}(X)$ the empirical mean of X and $\widehat{V}(X)$ its empirical variance.

Finally, the maximum likelihood estimate (MLE) also enables to determine the variance of the residual model. To reduce the number of full model parameters to be estimated and thus ease minimization of the log-likelihood, it is decided here to assign to each individual and measure the same covariance matrix of the discrepancy term (this assumption can be changed if deemed necessary):

$$\forall i \in [\![1, n]\!] \ \boldsymbol{\Omega}_i = \boldsymbol{\Omega} \ .$$

All the tools necessary for the calibration have been now described, it is possible to proceed to the calibration phase in the next Section. All results are shown with 3 significant digits. Figures are shown only for the repetition corresponding to the median of the KL divergence.

4.5 Calibration with synthetic data

This section is dedicated to the presentation of the calibration results in the mixed-effects framework with the sequential strategy using synthetic data. First, the same verification steps as in Chapter 3 are conducted to ensure the proper estimation of the individuals in Section 4.5.1, followed by the assessment of the opposite log-likelihood minimization in Section 4.5.2. Then, the calibration results are analyzed starting, with the tension experiments, to check that the joint distribution of S_{11}^0, ν_{12} and S_1^T is properly estimated in Section 4.5.3.1. The results of the second stage, with the compression experiments, are analyzed to assess the ability of the sequential procedure to estimate the full joint probability distribution in Section 4.5.3.2. Both orders are considered (tension to compression and then conversely), are compared to the direct calibration of the full joint distribution using design matrices in Section 4.5.4. The different steps are summarized in the flowchart depicted next.

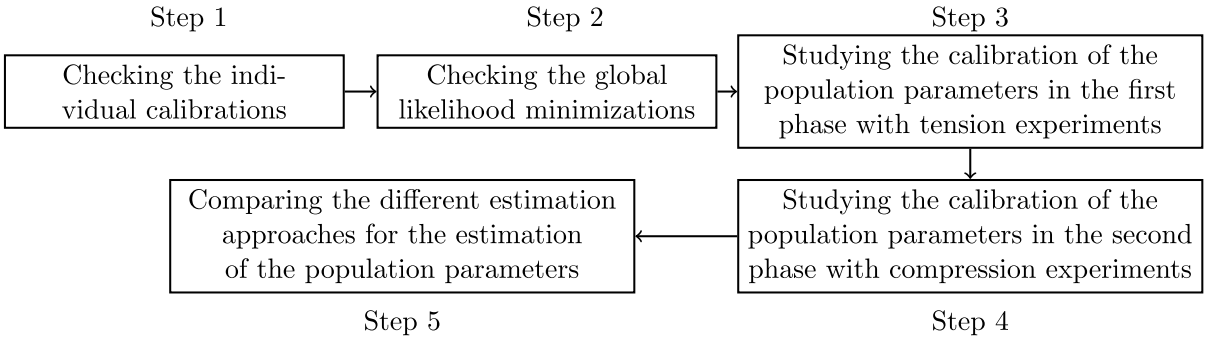


Figure 4.5: Flowchart of the different steps for the assessment of the mixed-effects calibration results.

4.5.1 Estimation of the individual parameters

The estimation of the population parameters is based on the individual parameters. Thus, it is necessary to validate that their estimation is carried out without major issues. To achieve this, we uniformly draw samples of $\boldsymbol{\Psi}$ (1,000) within the bounds defined in Table 4.3. For each of these samples, the individual parameters are estimated and the relative error is computed from Eq. (4.25).

Table 4.4: For the 8^{th} repetition and 1,000 proposals of full model parameters $\boldsymbol{\Psi}$ sampled between the bounds defined in 4.3, average relative errors $\mathbf{e}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_{50})$ in % on the individual parameters estimated with the Nelder-Mead algorithm.

	S_{11}^0	ν_{12}	S_1^T
Averaged relative error on the individual parameters in tension [%]	28.4	12.7	20.4

	S_{11}^0	ν_{12}	S_1^C
Averaged relative error on the individual parameters in compression [%]	22.0	9.05	19.6

In Table 4.4, it appears that the error on the individual parameters reach value comparable to those reached in Chapter 3. Still, these levels of errors are not negligible. One part of the explanation comes from the fact that the parameters are concentrated around their mean (the exact coefficient of variation is about 2% for S_{11}^0). Thus, for candidate means far from the exact values and small standard-deviations, the regularization term may prevent a proper estimation of the individual parameters. Furthermore, it should be noticed that S_{11}^0, S_1^T and S_1^C, involved in the prediction of both longitudinal and transverse outputs for the two types of experiments are flawed with higher errors. The trade-offs made between the transverse and longitudinal components may also explain these larger errors. In fact, it appears that the largest levels of errors are associated to $\boldsymbol{\Psi}$s with large opposite log-likelihoods, that will be not selected during the minimization. It therefore suggests that in practice, for the $\boldsymbol{\Psi}$s selected throughout the minimization process, the individual parameters are in fact properly estimated. This remain to be confirmed during the study of the results.

4.5.2 Minimization of the opposite log-likelihood

Before proceeding to the analysis, it is necessary to assess the minimization of the opposite log-likelihood by checking the history of the objective function value throughout the optimization, depicted in Figure 4.6 for all repetitions and for the traction and compression experiments.

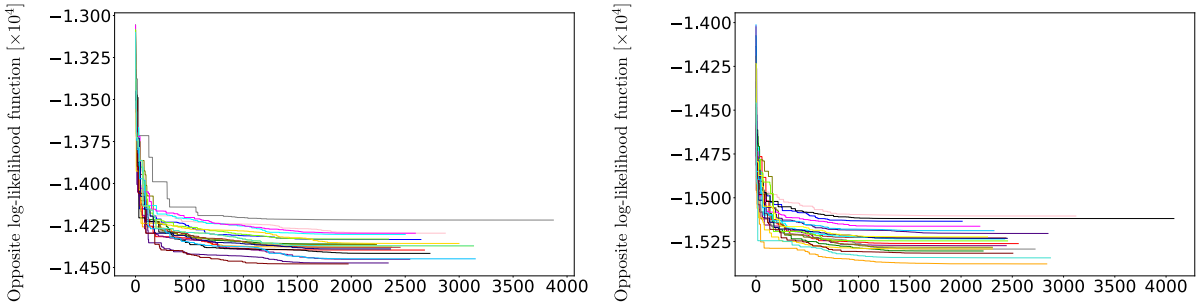


Figure 4.6: For all repetitions, best value of the objective function values for the tension experiments (on the left) and the compression experiments (on the right) with respect to the number of likelihood evaluations. The values of the first iterations are discarded for the sake of clarity.

In Figure 4.6, it can be seen that for all repetitions, the minimizations exhibit a proper convergence behavior. To finish this assessment, the variability of the calibration can be examined. To carry out this analysis, the minimization of the likelihood is performed for different initialization points that respectively correspond to $\nu \times \mathbf{1}_{\mathbb{R}^9}$ with $\nu \in \{0.15, 0.40, 0.65, 0.90\}$ (we limit the inquiry to 4 starts because of computational costs). Then, the relative difference δ between the values of the population parameters for the best calibration $\widehat{\boldsymbol{\Psi}}_{\text{best}}$ (in terms of likelihood) and the three other $\widehat{\boldsymbol{\Psi}}$s, defined as $\delta(\widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Psi}}_{\text{best}}) := 100 \times \frac{|\widehat{\boldsymbol{\Psi}} - \widehat{\boldsymbol{\Psi}}_{\text{best}}|}{\widehat{\boldsymbol{\Psi}}_{\text{best}}}$, are computed, averaged over the calibrations that do not refer to the best and reported for the 8^{th} repetition in Tables 4.5 and 4.6.

Table 4.5: Averaged relative difference δ in % of the population parameters for different initializations for the 8$^{\text{th}}$ repetition of the calibration process using the traction experiments. The random seed remains the same for all calibrations.

Population parameter	$\mu_{S^0_{11}}$	$\mu_{\nu_{12}}$	$\mu_{S^T_1}$	$\text{sd}(S^0_{11})$	$\text{sd}(\nu_{12})$	$\text{sd}(S^T_1)$
Averaged relative difference on the marginal parameters [%]	0.0412	0.00969	0.0.0409	0.292	1.74	0.454
			$\rho(S^T_{11}, S^0_{11})$	$\rho(\nu_{12}, S^T_{11})$	$\rho(\nu_{12}, S^0_{11})$	
Averaged relative difference on the correlations [%]			0.435	0.433	0.571	

Table 4.6: Averaged relative difference δ in % of the population parameters for different initializations for the 8$^{\text{th}}$ repetition of the calibration process using the compression experiments. The random seed remains the same for all calibrations.

Population parameter	$\mu_{S^0_{11}}$	$\mu_{\nu_{12}}$	$\mu_{S^C_1}$	$\text{sd}(S^0_{11})$	$\text{sd}(\nu_{12})$	$\text{sd}(S^C_1)$
Averaged relative difference on the marginal parameters [%]	0.0681	0.0478	0.0123	0.437	1.01	0.708
			$\rho(S^C_{11}, S^0_{11})$	$\rho(\nu_{12}, S^C_{11})$	$\rho(\nu_{12}, S^0_{11})$	
Averaged relative difference on the correlations [%]			2.15	0.338	0.317	

In Tables 4.5 and 4.6, it follows that except for the standard-deviation of S^0_{11}, the calibration seems to provide approximately the same estimation of the population parameters as the relative differences remain below 3%. This level of variability can be considered as low and does not prevent a proper estimation of the other parameters, and we can proceed now to the analysis of the calibration results.

4.5.3 First application of the sequential strategy

In this section, the detailed analyses from a first application of the sequential strategy starting with the tension experiments and ending with the compression experiments are presented.

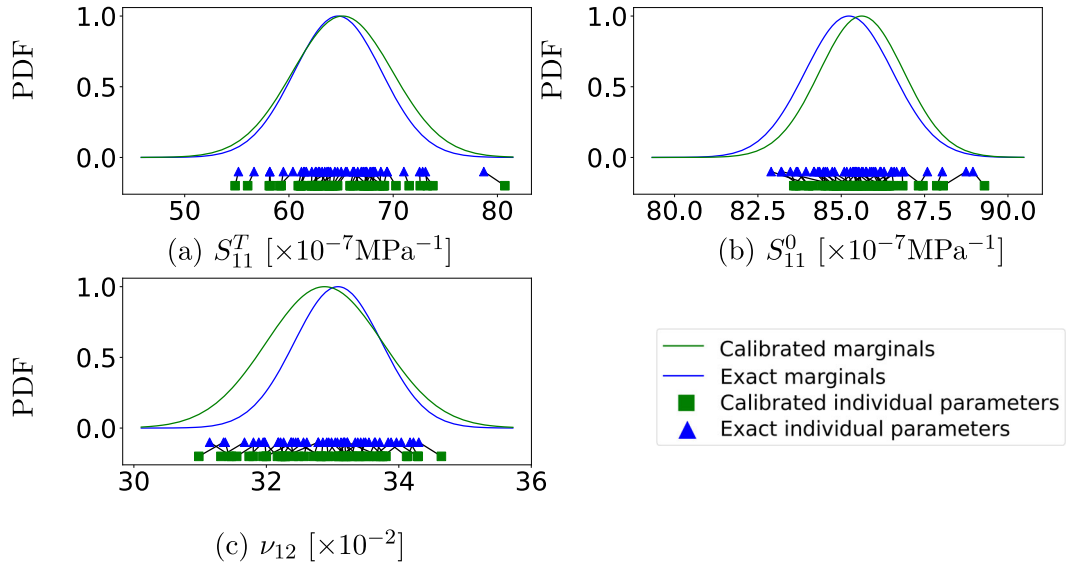
4.5.3.1 Calibration of the joint distribution of S^0_{11}, S^T_1 and ν_{12}

Given the modeling assumptions described in Sections 4.4.2 and 4.4.3.1, there are $p = 11$ parameters to be calibrated in this first step corresponding to the tension/compression step: 3 mean parameters, 3 standard-deviation parameters, 3 correlation parameters and 2 parameters for the residual model. The calibration of the dependence structure is achieved implementing a spherical parameterization of the Cholesky decomposition of the covariance matrix [Pinheiro and Bates, 1996]. This technique has multiple advantages, notably the guarantee to provide a positive-definite matrix for $\boldsymbol{\Sigma}$ without adding constraints to the optimization algorithm. The angles that are used to estimate the correlations are bounded between 10^{-5} and $(1 - 10^{-5})\pi$ (see Section 3.4.4 in Chapter 3). Another interest is that the optimization variables are easy to bound, contrary to those that appear in the Cholesky decomposition, which is helpful when defining the trust regions. To start the analysis result, the estimation of the calibrated means, standard-deviations and correlations reported in Table 4.7 can be discussed.

Table 4.7: Averaged calibrated means, standard-deviations and correlations $\widehat{\mathbf{\Pi}}_{1,\text{mean}}$ and averaged errors $\mathcal{E}_{\text{mean}}(\widehat{\mathbf{\Pi}}_1)$ in % over 20 repetitions of the calibration process with samples of 50 individuals each. The coefficients of variation COV in % are indicated between brackets.

	S_{11}^0 [$\times 10^{-7}$MPa^{-1}]	ν_{12} [$\times 10^{-2}$]	S_1^T [$\times 10^{-7}$MPa^{-1}]
Exact Means	85.2	33.1	64.7
Averaged calibrated means	85.3 (0.305)	33.1 (0.362)	64.8 (0.872)
Averaged errors on calibrated means [%]	0.249 (73.1)	0.306 (64.9)	0.738 (68.7)
Exact standard-deviations	1.31	0.660	4.19
Averaged calibrated standard deviations	1.19 (21.5)	0.672 (11.7)	4.08 (9.33)
Averaged errors on calibrated standard-deviations [%]	18.1 (65.9)	8.31 (105)	7.82 (68.5)
Correlations	$\rho(S_{11}^0, \nu_{12})$	$\rho(S_{11}^0, S_1^T)$	$\rho(S_1^T, \nu_{12})$
Exact correlations	-0.741	-0.500	0.809
Averaged calibrated correlations	-0.676 (23.4)	-0.450 (24.8)	0.851 (16.3)
Averaged errors on calibrated correlations [%]	17.9 (81.4)	18.8 (83.0)	15.9 (53.3)

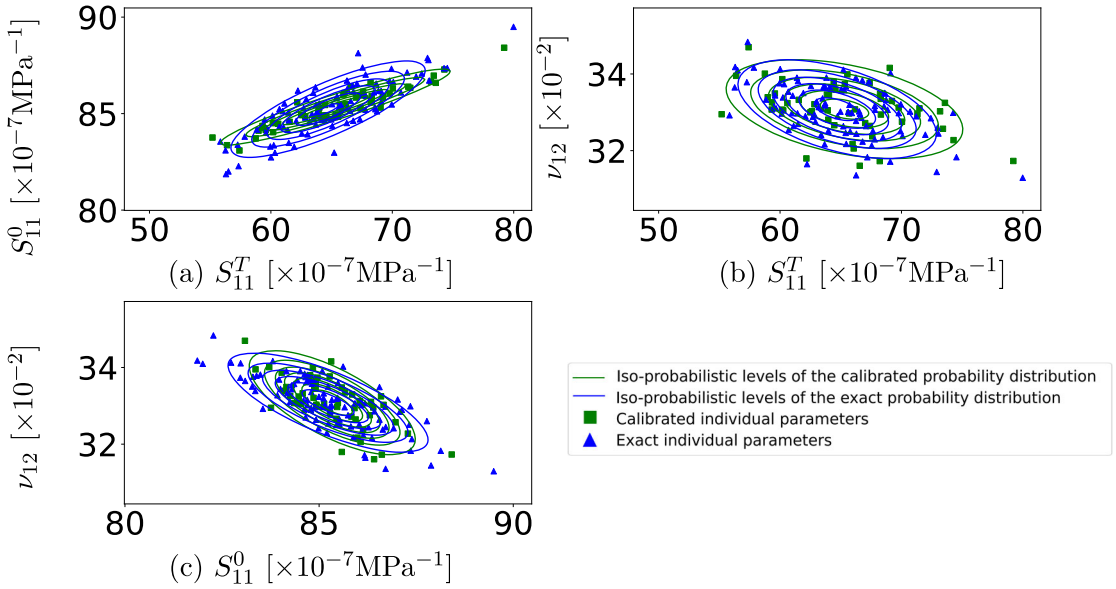
These results indicate that all the mean parameters are properly calibrated as the average relative error remains below 1% in average over all the repetitions. The standard-deviation parameters and the correlations are not estimated with the same accuracy, which can be understood as they require more samples to converge to their exact values. Still, the standard-deviations of both S_1^T and ν_{12} similarly to the correlation between them are estimated with level of errors (at most 15% in average) similar to those that could be found in Chapter 3. It is interesting to notice that even if 50 repetitions of experiments can be considered as a lot in a mechanical context, it remains relatively few in a statistical context as some repetitions reach high levels of error for the population parameters, as indicate the COVs. Remark that all population parameters linked to S_{11}^0 except its mean exhibit higher level of errors than the others. This suggests a poorer calibration of the corresponding individual parameters. To explain this point, remember that S_{11}^0 mainly expresses at the beginning of the strain-stress curve, where the residuals have the lowest values, while S_1^T expresses at the end of the experiment, for the highest residual values. Thus, function $g_i(\cdot)$ defined in Eq. (4.9) is necessarily more sensitive to the latter parameter than to S_{11}^0 that can be flawed with more error. Studying the correlations, the higher (in absolute value) the exact correlation, the more accurate the estimation. This is not surprising if remembering that correlations can be understood as constraints between the model parameters. The higher (in absolute value) it is, the more it coerces the individual parameters to follow a given pattern which makes the likelihood function sensitive to it, allowing to provide an accurate estimation. Globally, the calibration gives estimation of the joint distribution of S_{11}^0, S_1^T and ν_{12} that can be considered as rather accurate provided the difficulties coming from the concentration of the model parameters around their mean. The calibration of the marginals and the dependencies for the 8$^{\text{th}}$ repetition are depicted in Figure 4.8.



(a) S_{11}^T [$\times 10^{-7}$MPa^{-1}] (b) S_{11}^0 [$\times 10^{-7}$MPa^{-1}]

(c) ν_{12} [$\times 10^{-2}$]

Figure 4.7: Exact and calibrated marginals along the S_{11}^T axis (a), the S_{11}^0 axis (b) and the ν_{12} (c) axis for the 8$^{\text{th}}$ repetition. Exact and calibrated individual parameters are also plotted.



(a) S_{11}^T [$\times 10^{-7}$MPa^{-1}] (b) S_{11}^T [$\times 10^{-7}$MPa^{-1}]

(c) S_{11}^0 [$\times 10^{-7}$MPa^{-1}]

Figure 4.8: Exact and calibrated dependencies between S_{11}^T and S_{11}^0 (a), the S_{11}^T and ν_{12} (b) and ν_{12} and S_{11}^0 (c) for the 8$^{\text{th}}$ repetition. Exact and calibrated individual parameters are also plotted.

To check the consistency of the calibration, it is mandatory to study the calibration of the individual parameters as they influence the determination of the population parameters. Table 4.8 reports the error on the individual parameters as well as the error on strain outputs. It also contains the measure of the spread of each model parameter labeled ς_{exact} (instead of the empirical COV), that is to say $\varsigma_{\text{exact}} := 100 \times \frac{\sqrt{\text{diag}(\boldsymbol{\Sigma}_{\text{exact}})}}{\boldsymbol{\mu}_{\text{exact}}}$ with $\boldsymbol{\mu}_{\text{exact}}$ and $\boldsymbol{\Sigma}_{\text{exact}}$ the exact mean vector and covariance matrix. It measures how much each parameter is concentrated around its mean values. In fact, it is a coefficient of variation that is here renamed, to prevent confusions with the empirical coefficient of the averaged relative errors (denoted COV), that also measures the concentration around their mean value for several quantities that are different from $\boldsymbol{\theta}$.

Table 4.8: Relative errors in % on the estimation of the individual parameters $\mathbf{e}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_{50})$ averaged over 20 repetitions of the calibration process with 50 individuals along with the spread ς_{exact}. The corresponding coefficients of variation in % are indicated between brackets.

	S_{11}^0	ν_{12}	S_1^T
Relative error [%]	0.616 (41.0)	0.613 (57.4)	1.6 (135)
Spread ς_{exact} [%]	1.53	2.00	6.48

Table 4.9: Error in model space $\mathbf{D}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_{50})$ for each strain component.

Strain component	ε_{11} [-]	ε_{22} [-]
Order of magnitude of the mean strain value	10^{-3}	10^{-3}
Distance in model space	8.67×10^{-9}	8.66×10^{-10}

In Tables 4.8 and 4.9, the individual parameters are, in average, always properly estimated with a maximum error of at most 1.6% for each parameter. This is a key feature as a poor calibration of the individual parameters would prevent a proper estimation of the joint distribution of the model parameters.
Furthermore, in model space, the square root of the model error reaches about 10^{-5} for both strain component, that can be considered low compared to the strain values that are above 10^{-3} for most of the experiments. This confirms the good calibration of the experiments in average for all repetitions. It has been stated earlier that the linear compliance was not estimated as accurately as it could be because the residual of the nonlinear part outweight those of the linear behavior monitored by S_{11}^0. Yet, this statement seems to be in contradiction with the level of error. Meanwhile, a comparison to the exact spread values ς_{exact} can bring some elements of justification. Indeed, this variable is the most concentrated around its mean value, especially compared to S_1^T. Thus, within $[\mu_{S_{11}^0} - 3\text{sd}(S_{11}^0), \mu_{S_{11}^0} + 3\text{sd}(S_{11}^0)]$, (the 99% credible interval for Gaussian distributed variables), most of the parameters will give the same model error, which makes them difficult to identify exactly. This effect has a small impact on the mean parameter because it is a position parameter and the same interval of linear compliances is always targeted. However, this gets different for the standard-deviation which indicates the dispersion of the model parameters and depends on the position of the individual parameters with respect to the mean value, making it more sensitive to estimation errors.
To terminate the assessment, a brief check of the residual can be conducted. For the 8^{th} repetition of the calibration, the raw residuals from the 6^{th} test, that is to say the difference between the noisy and predicted strains $\widehat{\mathbf{r}}_6 := \mathbf{y}_6 - \mathcal{F}(\mathbf{t}_6, \widehat{\boldsymbol{\theta}}_6)$ are displayed in Figure 4.9. The Gaussianity of the residuals is assessed by plotting their QQ-plots in Figure 4.10.

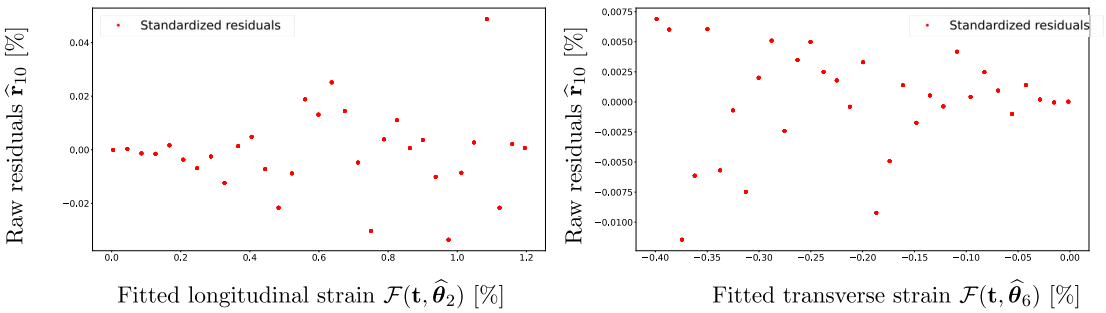


Figure 4.9: For the 8^{th} repetition, raw residuals for the 6^{th} for the longitudinal (on the left) and transverse component (on the right).

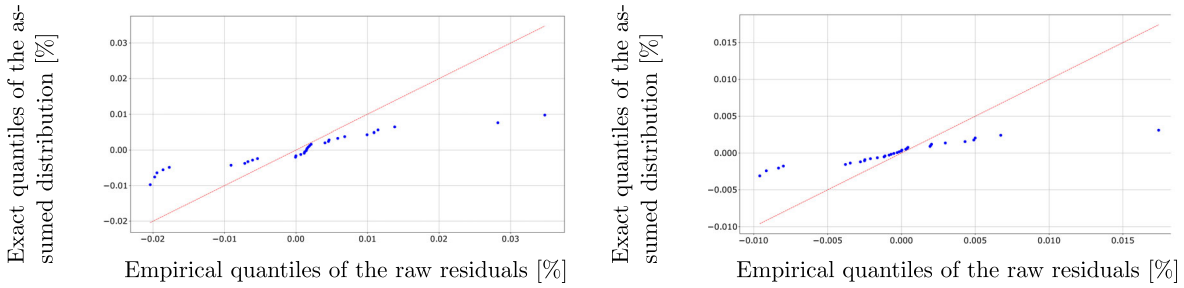


Figure 4.10: For the 8$^{\text{th}}$ repetition, QQplot of the residuals for the 6$^{\text{th}}$ individual for the longitudinal (on the left) and transverse component (on the right). In red, the first bisector on which data points should be located.

In Figure 4.9, it can be seen that the residuals increase with the stress values such as their variance, suggesting that they are heteroscedastic (which is indeed the case). However, this does not seem to prevent a satisfactory calibration of the model parameter distribution and may only delay the convergence of the minimization process. In addition, it can be noticed that the raw residuals refer to at most 5% of the longitudinal fitted strains and 4% of the transverse fitted strains confirming the proper calibration of the individuals illustrated in Tables 4.8 and 4.9. In fact, as in the previous chapter, the assumption of homoscedastic residuals seems to be sufficient to achieve the calibration of the population parameters. Similarly, the Gaussian assumption of the residuals is not confirmed. Though the estimated population parameters are consistent with the exact values, their estimation may be improved by considering more complicated residual model (*e.g.*, with heteroscedasticity). However, this would represent an increase of the complexity of the statistical modeling, potentially leading to an increase of the complexity of the resolution method. In particular, this could decrease the quality of the identification of the model of the individuals, flawing the calibration of the population parameters. This subject is a major concern and deserves to be studied further in a dedicated work. Thus, despite of the limitations highlighted here, these different indicators allow to validate the calibration of joint distribution of S_{11}^0, S_1^T and ν_{12} either at the population or individual levels, and the first stage of the methodology can be considered as completed.

4.5.3.2 Calibration of the joint distribution of S_{11}^0, S_1^C and ν_{12}

In the second stage of the methodology, the objective is to identify the joint distribution of S_{11}^0, S_1^C and ν_{12}. Furthermore, the parameters of the joint distribution of ν_{12} and S_{11}^0: $\mu_{\nu_{12}}, \mu_{S_{11}^0}, \mathbb{V}(\nu_{12}), \mathbb{V}(S_{11}^0)$ and $\rho(\nu_{12}, S_{11}^0)$, already estimated in the previous step, will be limited to a trust region corresponding to $\mathcal{R}_1 = [0.8\mathbf{\Pi}_1, 1.2\mathbf{\Pi}_1]$. Considering 20% of variation both enables a small exploration around the calibrated value of $\mathbf{\Pi}_1$ and to remain concentrated around them. To bound the correlation parameter, $\rho(\nu_{12}, S_{11}^0)$, a spherical decomposition of the covariance matrix is used [Pinheiro and Bates, 1996]. The number of population parameters to estimate remains the same (11), with six of them limited to $\mathcal{R}_1$.
In this step, we should ensure that both the extracted covariance matrix (that is to say the blue matrix in Eq. (4.18)) and the correlation matrix of the complete joint distribution are consistent. Constraints can be implemented but it is possible to avoid them. In the first step, the columns (equivalently the rows) of the correlation matrix have been ordered as follows: S_{11}^T, S_{11}^0 and ν_{12}. To generalize the matrix from the first to the second step, we add in this step a fourth row (so also a fourth column to ensure that we keep a square matrix) that gives us the complete correlation whose rows (columns) are ordered in this order: S_{11}^T, S_{11}^0, ν_{12} and S_{11}^C. Then, the column and the row refering to S_1^T are deleted (not estimated). This process is illustrated in Figure 4.11.

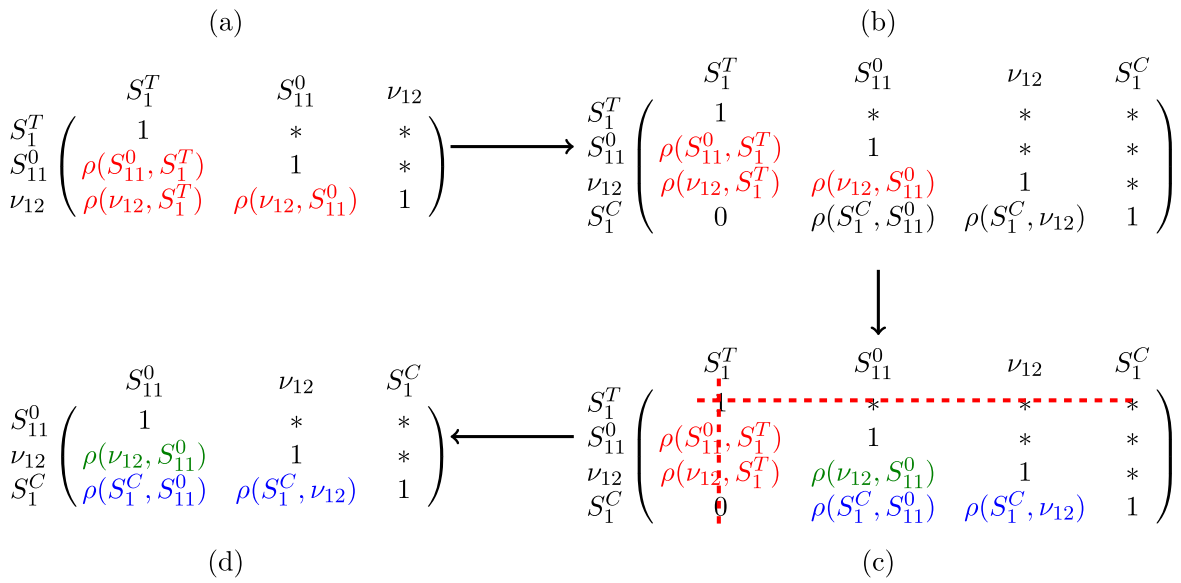


Figure 4.11: Construction of the correlation matrix in the second step. Stage (a) refers to the calibration with the tension experiments, and b)c) and d) to the calibration with the compression experiments. The correlations estimated with the traction experiments are highlighted in red, the one estimated twice in green and those calibrated with the compression experiments in blue.

Figure 4.11 details how the correlation matrix is assembled in the second step of the methodology (top left). a) First, the correlation matrix estimated in the first step is assembled. b) It is plugged within the 4×4 full correlation matrix that the correlations linked to S_1^C. c) Then, the row and column associated to S_1^T are deleted which provides the final correlation matrix d). In this case, the described procedure is the only one that is constraint-free. The calibrated means, standard-deviations and correlations are given in Table 4.10.

Table 4.10: Averaged calibrated means, standard-deviations and correlations $\widehat{\mathbf{\Pi}}_{2,\text{mean}}$ and averaged errors $\mathcal{E}_{\text{mean}}(\widehat{\mathbf{\Pi}}_2)$ in % over 20 repetitions of the calibration process with samples of 50 individuals each. The coefficients of variation COV in % are indicated between brackets. It is continued next page.

	S_{11}^0 [$\times 10^{-7}$MPa^{-1}]	ν_{12} [$\times 10^{-2}$]	S_1^C [$\times 10^{-7}$MPa^{-1}]
Exact Means	85.2	33.1	155.2
Averaged calibrated means	85.4 (0.424)	33.1 (0.339)	155.4 (1.89)
Averaged errors on calibrated means [%]	0.376 (76.5)	0.273 (74.6)	1.52 (74.0)
Exact standard-deviations	1.31	0.660	17.3
Averaged calibrated standard deviations	1.01 (28.6)	0.669 (10.5)	17.9 (7.93)
Averaged errors on calibrated standard-deviations [%]	26.3 (67.5)	9.41 (54.8)	6.68 (84.9)
Correlations	$\rho(S_{11}^0, \nu_{12})$	$\rho(S_1^C, \nu_{12})$	$\rho(S_{11}^0, S_1^C)$
Exact correlations	−0.741	0.729	−0.414
Averaged calibrated correlations	−0.726 (18.6)	0.713 (6.40)	−0.354 (56.0)
Averaged errors on calibrated correlations [%]	14.7 (74.6)	4.91 (90.4)	41.1 (69.0)

First, one can notice that the distribution parameters involving S_1^C (not calibrated in step 1) are well estimated. Indeed, the average relative error on the marginal parameters is, over all the repetitions, below 10%. The correlation coefficients show larger level of errors, but it can be understood as it is rather weak (-0.41) which adds to the limited number of individuals. Otherwise, the calibration of the population parameters that involve S_1^C can be considered as satisfactory, provided that this parameter is the most dispersed around its mean value (its exact coefficient of variation is about 10% which makes it more subjected to estimation errors.)
It is now possible to focus on the joint distribution of S_{11}^0 and ν_{12}. From Table 4.10, it seems that the estimation on this joint distribution is downgraded between the two steps. Indeed, except for the mean parameters, the two standard-deviations appear to be flawed from their initial estimation (Table 4.7), in particular for the standard-deviation of S_{11}^0 (the averaged error increases by 30%). This downgrade can be noticed both the averaged values or averaged errors on the population parameters. The correlation between S_{11}^0 and ν_{12} does not seem to be much affected and even exhibits a small improvement.
Rather than focusing independently on each parameter, a more suited way to quantify this alleged degradation is the Kullback-Leibler divergence between the exact joint distribution of S_{11}^0, ν_{12} ($f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}_{\text{exact}}}(S_{11}^0, \nu_{12})$) and the calibrated distribution ($f_{\boldsymbol{\Theta}|\widehat{\boldsymbol{\Pi}}}(S_{11}^0, \nu_{12})$) in the two stages as depict Figure 4.12 for the 20 repetitions.

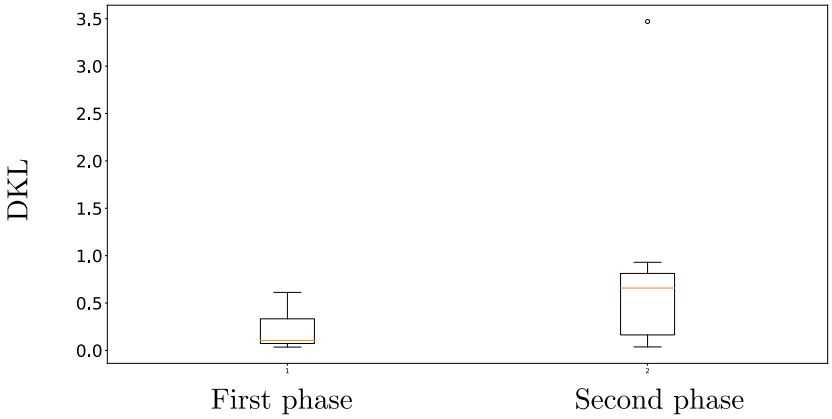
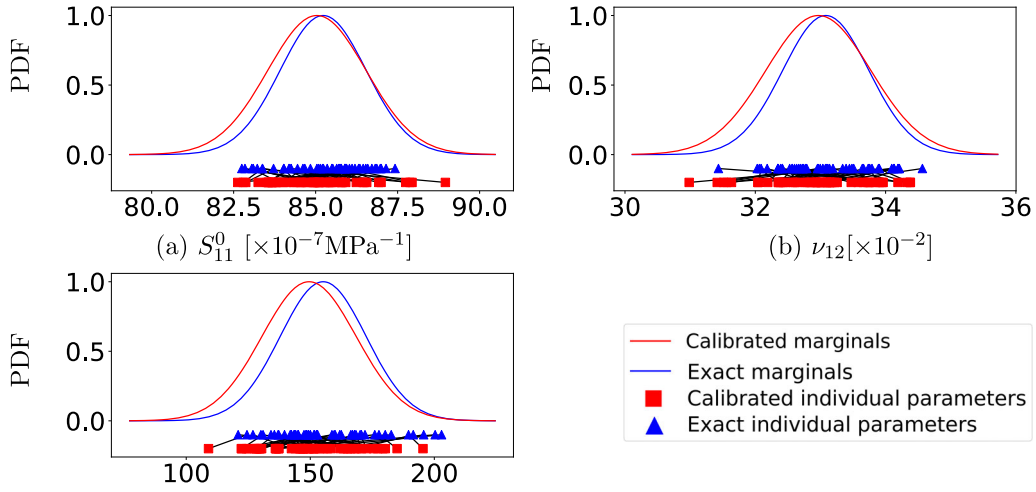


Figure 4.12: KL divergences (DKL) of the calibrated and exact joint distribution S_{11}^0 and ν_{12} for the 20 repetitions in the two stages. Remember that the KL divergence expresses a notion of discrepancy between two distributions and the lower is its value, the more accurate the calibration of $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ is.

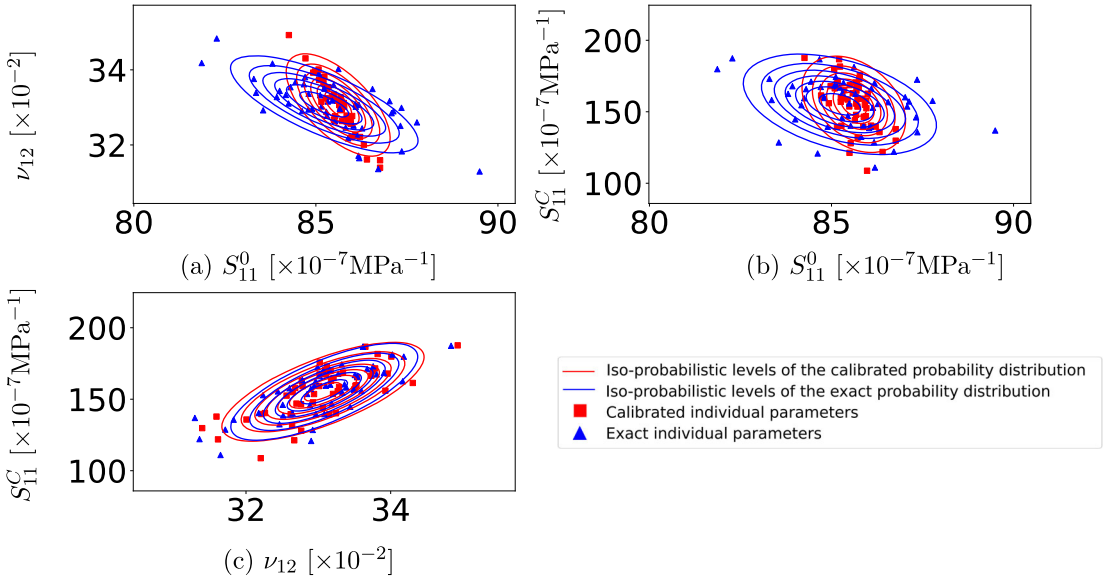
In Figure 4.12, the deterioration of the joint distribution of S_{11}^0 and ν_{12} suggested by Table 4.10 seems to be confirmed. For instance, the median of the KL divergences increases (multiplied by 3), and the ranges of variation of the KL divergences are also increased by two (without consideration of the outliers). This seems mainly due to poorer estimation in the compression phase of the standard-deviation of S_{11}^0. Consequently, a deterioration of the estimation of joint distribution of S_{11}^0 and ν_{12} can be observed in the second phase. Nevertheless, this deterioration remains limited for both the relative errors and the KL divergences. In particular, even if they are larger in the second phase than in the first phase, the KL divergences have the same orders of magnitude. Indeed, except for one repetition, the KL divergences range from 0 to 0.7 in the first step and from 0 to 1 in the second step, which can be considered as a clear but small downgrade. This does not necessarily invalidate the sequential method, that would require a comparison with the direct estimation method proposed in Section 4.5.4 but rather hints that the calibration is harder to perform using the compression experiments instead of tension experiments. One possible explanation could be a poorer calibration of the individual linear compliances S_{11}^0 (to be investigated later). The calibration of the marginals for the 8^{th} repetition of the calibration process are depicted in Figure 4.13 and the dependencies in Figure 4.14.



(a) S_{11}^0 [$\times 10^{-7}$MPa^{-1}] (b) ν_{12}[$\times 10^{-2}$]

(c) S_{11}^C [$\times 10^{-7}$MPa^{-1}]

Figure 4.13: Exact and calibrated marginals along the S_{11}^0 axis (a), the ν_{12} axis (b) and the S_{11}^0 axis (c) for the 8$^{\text{th}}$ repetition of the calibration process. Exact and calibrated individual parameters are also plotted.



(a) S_{11}^0 [$\times 10^{-7}$MPa^{-1}] (b) S_{11}^0 [$\times 10^{-7}$MPa^{-1}]

(c) ν_{12} [$\times 10^{-2}$]

Figure 4.14: Exact and calibrated dependencies between ν_{12} and S_{11}^0 (a), the S_{11}^C and S_{11}^0 (b) and ν_{12} and S_{11}^C (a) for the 8$^{\text{th}}$ repetition of the calibration process. Exact and calibrated individual parameters are also plotted.

To conclude the analysis, it remains to check the calibration of the individuals, either in the parameters and model space. The suited indicators are displayed in Tables 4.11 and 4.12.

Table 4.11: Relative errors on the estimation of the individual parameters $\mathbf{e}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_{50})$ in % averaged over 20 repetitions of the calibration process with 50 individuals along with the spread ς_{exact}. The corresponding coefficients of variation in % are indicated between brackets.

	S_{11}^0 [$\times 10^{-7}$MPa^{-1}]	ν_{12} [$\times 10^{-2}$]	S_1^C [$\times 10^{-7}$MPa^{-1}]
Relative error [%]	1.50 (13.0)	2.22 (9.83)	13.1 (10.1)
Spread ς_{exact} [%]	1.53	2.00	10.0

Table 4.12: Error in model space $\mathbf{D}(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_{50})$ for each strain component.

Strain component	ε_{11} [-]	ε_{22} [-]
Order of magnitude of the mean strain value	10^{-3}	10^{-3}
Distance in model space	6.36×10^{-8}	9.00×10^{-9}

From Tables 4.11 and 4.12, the same remarks made on Tables 4.8 and 4.9 arise. Indeed, in average, for all repetitions, all experiments seem properly calibrated either in model or parameters space. However, with respect to their exact spread, the nonlinear compliance can be considered as better calibrated than the others. The only difference is that the elastic linear parameters S_{11}^0, ν_{12} are flawed with higher errors, partly explaining the poorer calibration of their joint distribution highlighted above. This could come from the nonlinearity of the strain-stress curve more pronounced in compression than in tension, making the sum of the residuals of the nonlinear part greater than those from the linear part, here reduced to few points, but this remains to be confirmed by a detailed study.

The results from this section and the previous one show that by conducting the calibration starting with the tension experiments and then continuing with the compression experiments, it is possible to estimate accurately the full joint distribution of the longitudinal nonlinear elasticity, despite of the numerical complexities coming from the concentration of the model parameters around their means. Still, the sequential calibration has been performed in an arbitrarily chosen order and there is no particular reason to favor this one to others. Thus, it is necessary to check to what the extent the choice of the order and more generally the calibration strategy influences the calibration results. In the next section, the sequential strategy is achieved in the reverse order to check in particular whether this choice downgrades the estimation of the model parameters distribution, before being compared to the method that implements design matrices.

4.5.4 Impact of the order of calibration

In this section, the objective is to present the results provided by the implementation of the sequential strategy in the reverse order: compression first, then traction. All the optimization settings for the sequential scheme are kept the same than in Section 4.5.3. The main issue here is to know whether one of the orders provides better results (in terms of errors). To investigate this point, the result of main interest is directly studied, *i.e.*, how does evolve the calibration of the joint distribution of S_{11}^0, ν_{12} ? In particular, is the estimation downgraded or improved between the two stages ? To answer this question, the KL divergence between the exact joint distribution of S_{11}^0, ν_{12} ($f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}_{\text{exact}}}(S_{11}^0, \nu_{12})$) and the calibrated distribution ($f_{\boldsymbol{\Theta}|\widehat{\boldsymbol{\Pi}}}(S_{11}^0, \nu_{12})$) in the two stages can be estimated.

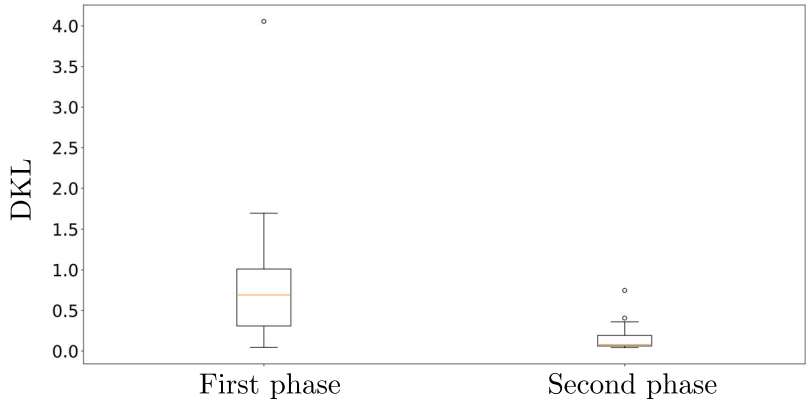


Figure 4.15: KL divergences (DKL) of the calibrated and exact joint distribution S_{11}^0 and ν_{12} for the 20 repetitions in the two stages.

Here, contrary to the previous case, there is a clear improvement of the calibration results between the two phases. Indeed, the median of the KL divergences is almost divided by 7. This is interesting because, here, the tension phase corrected the errors made in the compression phase that appears harder to conduct as shown by all the indicators studied so far. In fact, the observed decrease mainly comes from the fact that, with the compression experiments, the estimation of the population parameters that concern the joint distribution of S_{11}^0 and ν_{12} shows higher levels of error compared the to estimate of the population parameters from the tension experiments, as it arises from Tables 4.7 and 4.10. Thus, as the step involving the compression experiments that provides the poorest results was conducted first, at the end of the sequential process, the calibrated distribution is estimated with better accuracy contrary to when compression experiments are considered at the end of the calibration strategy.
It is also worth noticing that, with the compression experiments, the joint distribution of the S_{11}^0 and ν_{12} is estimated with increased accuracy when its calibration is carried out in the second phase rather than in the first phase, as show the KL divergences in Figures 4.12 and 4.15. Indeed, though their medians are comparable in both situations, the range of variation of the KL divergences (without consideration of the outliers) in the second phase of the calibration process is about half of their range of variation in the first phase of the calibration process when starting with the compression experiments. This illustrates the interest to use a sequential scheme. Indeed, the MLE calibrated in the first phase (conducted with the tension experiments and described in Section 4.6.3) is used to define a trust region, which reduces the size of the search space by focusing the optimization within a region that concentrates proposals of population parameters that seems to be close to the value of MLE of the second phase. This eases the optimization and seems to limit the deterioration of the population parameters estimates, contrary to the situation in which the population parameters are learnt from scratch (that is to say when the compression experiments are used first). Still, this assumption would demand to be consolidated with uncertainty quantification techniques and confirmed by appropriate convergence studies
It can be interesting to know whether this is a global improvement (*i.e.*, the estimation of all parameters was improved) or if it is concentrated on one specific parameter. To study this particular point, the averaged error on the population parameters of the joint distribution of S_{11}^0 and ν_{12} whose values are reported in Tables 4.13 and 4.14 can be investigated.

From these tables, it follows that the estimation of all the population parameters is slightly improved from the first to the second phase except $\mathrm{sd}(S_{11}^0)$, whose averaged estimation error is reduced by about 30%. This mainly comes from the fact that the linear compliances are identified with better accuracy with the tension experiments than with the compression experiments, as show Tables 4.8 and 4.11. Still, it can be noticed that the calibration of the standard-deviation of ν_{12} and of the correlation is not improved as for the other parameters given that the coefficient of variation of their corresponding averaged error is increased, indicating that this improvement is not uniform among the repetitions. However, globally, with this particular model, provided the modeling assumptions and the values chosen for population parameters, it seems better to identify first the joint distribution of the population parameters in compression, before a refinement with the tension experiments. Of course, this conclusion does not apply in all cases and is only specific to our model of interest, parameters value and modeling choices.

To terminate this comparison, let's see how this method performs with respect to the application of standard mixed-effects presented in the next section.

Table 4.13: Averaged marginal parameters ($\widehat{\mathbf{\Pi}}_{1,\text{mean}}$ from the first phase and $\widehat{\mathbf{\Pi}}_{2,\text{mean}}$ from the second phase), alongside the averaged error ($\mathcal{E}_{\text{mean}}(\widehat{\mathbf{\Pi}}_1)$ and $\mathcal{E}_{\text{mean}}(\widehat{\mathbf{\Pi}}_2)$ in %) over 20 repetitions of the calibration process with samples of 50 individuals each. The first two lines refer to the error made on the mean parameters in both phases, and the two left to the standard-deviation parameters estimated in the second phase in both phases. The coefficients of variation COV in % are indicated between brackets.

	S_{11}^0 [$\times 10^{-7}$MPa^{-1}] (1$^{\text{st}}$ phase)	ν_{12} [$\times 10^{-2}$] (1$^{\text{st}}$ phase)
Exact Means	85.2	33.1
Averaged calibrated means	85.4 (0.438)	33.1 (0.343)
Averaged errors on calibrated means [%]	0.364 (78.9)	0.277 (73.4)
	S_{11}^0 [$\times 10^{-7}$MPa^{-1}] (2$^{\text{nd}}$ phase)	ν_{12} [$\times 10^{-2}$] (2$^{\text{nd}}$ phase)
Exact means	85.2	33.1
Averaged calibrated means	85.3 (0.306)	33.1 (0.366)
Averaged errors on calibrated means [%]	0.251 (73.1)	0.309 (64.8)
	S_{11}^0 [$\times 10^{-7}$MPa^{-1}] (1$^{\text{st}}$ phase)	ν_{12} [$\times 10^{-2}$] (1$^{\text{st}}$ phase)
Exact standard-deviations	1.31	0.660
Averaged calibrated-standard-deviations	1.07 (30.1)	0.675 (10.5)
Averaged errors on calibrated standard-deviations [%]	25.5 (73.9)	9.44 (59.5)
	S_{11}^0 [$\times 10^{-7}$MPa^{-1}] (2$^{\text{nd}}$ phase)	ν_{12} [$\times 10^{-2}$] (2$^{\text{nd}}$ phase)
Exact standard-deviations	1.31	0.660
Averaged calibrated-standard-deviations	1.20 (21.1)	0.669 (10.7)
Averaged errors on calibrated standard-deviations [%]	17.7 (64.7)	7.48 (108)

Table 4.14: Averaged correlation parameters $\widehat{\rho}(S_{11}^0, \nu_{12})_{\text{mean},1/2}$ in both phases and averaged error $\mathcal{E}_{\text{mean}}(\widehat{\rho}(S_{11}^0, \nu_{12}))_{\text{mean},1/2}$ in % over 20 repetitions of the calibration process with samples of 50 individuals each. The first column refers to the error made on the population parameters in the second stage of the sequential methodology (with the compression experiment on the left and with the tension experiment on the right). The coefficients of variation COV in % are indicated between brackets.

	$\rho(S_{11}^0, \nu_{12})$ (1$^{\text{st}}$ phase)	$\rho(S_{11}^0, \nu_{12})$ (2$^{\text{nd}}$ phase)
Exact correlation	-0.741	-0.741
Averaged calibrated correlations	-0.771 (20.8)	-0.715 (18.5)
Averaged errors on calibrated correlations [%]	18.2 (68.2)	13.8 (85.8)

4.5.5 Comparison of the sequential strategy and standard mixed-effects

So far, the results provided by a sequential strategy have only been discussed. However, in this particular case, it is possible to use classical mixed-effects to estimate, with the help of design matrices, the 13 parameters of the joint distribution of S_1^T, S_1^C, S_{11}^0 and ν_{12} (recall that $\text{Cov}(S_1^T, S_1^C)$ cannot be estimated with the actual data). Furthermore, with the two different kind of experiments considered here, the tension and compression experiments, a reasonable assumption is to consider different residual models for each kind of experiment. Provided that there are two different curves to identify (longitudinal and transverse measurements), that makes 4 parameters for the two residual models. In total, $p = 17$ parameters have to be estimated. This illustrates a first interest to use a sequential scheme as with such techniques, only $p = 11$ parameters have to be calibrated in the different phases. The three calibration techniques are studied here. They respectively stand for the calibrated distribution in the second phase of the first implementation of the sequential strategy (the one obtained after the calibration with the compression experiments) labeled $\mathcal{C}_1$, the calibrated distribution in the second phase of the second implementation (the one obtained after the calibration with the tension experiments) $\mathcal{C}_2$ and finally, the one that comes from the complete calibration the joint distribution of S_1^T, S_1^C, S_{11}^0 and ν_{12}, labeled $\mathcal{C}_3$. Several items can be investigated among which the error on all population parameters in the three cases, as reported in Table 4.15.

Table 4.15: Averaged error $\mathcal{E}_{\text{mean}}(\widehat{\mathbf{\Pi}})$ in % on the population parameters calibrated in the three cases over 20 repetitions of the calibration process with samples of 50 individuals each. For cases $\mathcal{C}_1$ and $\mathcal{C}_2$, the error on the parameters of the joint distribution of S_{11}^0 and ν_{12} correspond to those obtained after completion of the sequential scheme, that is to say after the second step. The coefficients of variation COV in % are indicated between brackets.

Population parameter	$\mu_{S_{11}^0}$	$\mu_{\nu_{12}}$	$\mu_{S_1^T}$	$\mu_{S_1^C}$
Averaged errors in case $\mathcal{C}_1$ [%]	0.376 (76.5)	0.273 (74.6)	0.738 (68.7)	1.52 (74.0)
Averaged errors in case $\mathcal{C}_2$ [%]	0.251 (73.1)	0.309 (64.8)	0.736 (66.7)	1.60 (72.2)
Averaged errors in case $\mathcal{C}_3$ [%]	0.259 (78.1)	0.281 (68.2)	0.683 (77.9)	1.61 (72.6)
Population parameter	$\text{sd}(S_{11}^0)$	$\text{sd}(\nu_{12})$	$\text{sd}(S_1^T)$	$\text{sd}(S_1^C)$
Averaged errors in case $\mathcal{C}_1$ [%]	26.3 (67.5)	9.41 (54.8)	7.82 (68.5)	6.68 (84.9)
Averaged errors in case $\mathcal{C}_2$ [%]	17.7 (64.7)	7.48 (108)	7.85 (67.4)	7.11 (75.7)
Averaged errors in case $\mathcal{C}_3$ [%]	18.1 (69.5)	7.96 (87.3)	7.90 (78.5)	7.43 (73.3)

Population parameter	$\rho(\nu_{12}, S_1^T)$	$\rho(\nu_{12}, S_1^C)$	$\rho(S_1^C, S_{11}^0)$	$\rho(S_1^T, S_{11}^0)$	$\rho(S_{11}^0, \nu_{12})$
Averaged errors in case $\mathcal{C}_1$ [%]	18.8 (83.0)	4.91 (90.4)	41.1 (69.0)	17.9 (81.4)	14.7 (74.6)
Averaged errors in case $\mathcal{C}_2$ [%]	26.7 (97.8)	5.44 (90.3)	53.9 (72.8)	16.9 (69.4)	13.8 (85.8)
Averaged errors in case $\mathcal{C}_3$ [%]	26.3 (111)	8.79 (168)	68.0 (121)	25.8 (76.9)	16.4 (165)

In Table 4.15, it is noticeable that the estimation of the marginals of the asymptotic compliance in tension and compression are slightly downgraded in case $\mathcal{C}_3$ compared to cases $\mathcal{C}_1$ and $\mathcal{C}_2$. This could come from the fact in case $\mathcal{C}_3$, all variables are considered together and may be more difficult to tune than when the population parameters corresponding to a single type of experiment are estimated. Still, the differences remain small and may also originate from numerical issues. Furthermore, it appears that, surprisingly, the calibration that considers both compression and tension does not necessarily provide the best results for the calibration of the joint distribution of S_{11}^0 and ν_{12}. Indeed, as these parameters express on both types of experiments with 50 individuals each (so 100 individuals in total), we could expect it to be

the best solution among the three possibilities given that this distribution is estimated with the highest number of individual possible. Obviously, it is not true in this particular case. In fact, it seems that the calibration process makes trade-offs between the tension and compression experiments, and provided that the linear compliances are identified with less accuracy on the compression than the tension experiments. The phenomenon of trade-off can be illustrated in the parameters space by depicting the PDFs of the elastic linear parameters, estimated with either the tension or compression experiments in Figure 4.16.

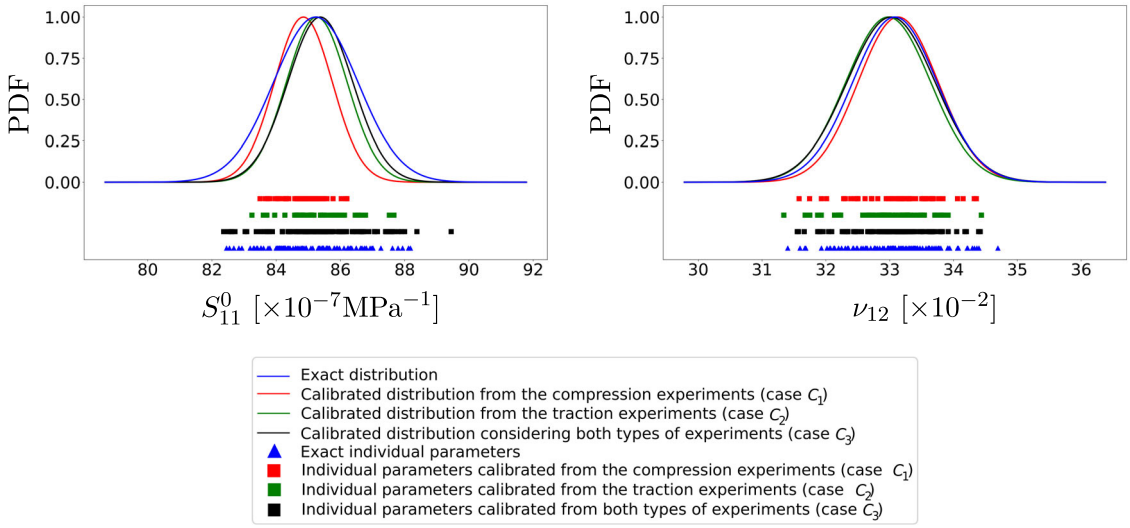


Figure 4.16: For the 8$^{\text{th}}$ repetition, PDF of S^0_{11} and ν_{12} and the corresponding individual parameters for the calibration from case $\mathcal{C}_1$ (with the compression experiments in the second phase), from case $\mathcal{C}_2$ (with the tension experiments in the second phase) and considering both types of experiments (case $\mathcal{C}_3$).

In Figure 4.16, on the left, it can be noticed that, if the calibration is carried out only from the compression experiments, because of the higher levels of error, the marginal parameters of S^0_{11} are estimated less accurately than when they are estimated from only the tension experiment or when both types of experiments are jointly considered, which is consistent with previous results. For this repetition, adding the compression experiments to the tension one slightly decreases the error on the marginal estimation. This seems to be linked to the level of errors on the individual parameters as discussed earlier. The calibration of the other marginals and correlations do not exhibit specific problems provided the limitations of the population parameters values and the model issues. The KL divergence is computed for two distributions: the joint distribution of S^0_{11} and ν_{12} on the one hand and on the other hand for the full joint distribution, and illustrated in Figure 4.17.

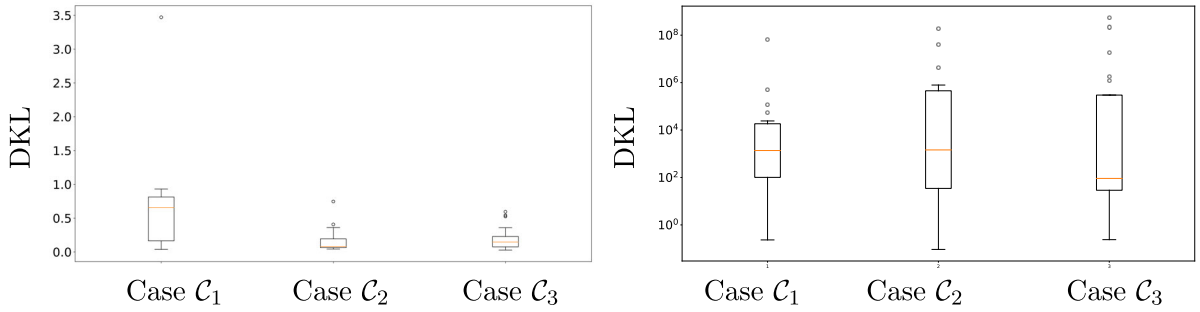


Figure 4.17: On the left, KL divergences (DKL) between the calibrated and exact joint distribution of S^0_{11} and ν_{12} and on the right, KL divergences between the calibrated and exact joint full distributions for the 20 repetitions and 3 cases.

In Figure 4.17, it is interesting to notice that the KL divergences are similar in terms of values and range of variation, either for the joint distribution of S_{11}^0 and ν_{12} or for the complete joint distribution. Still, it is possible to observe the KL divergences have significant range of variation (from 10^{-1} to over 10^8). This can be partly explained by the fact that the complete distribution is parameterized by 13 parameters (against 5 for the joint distribution of S_{11}^0 and ν_{12}) whose 2 correlations exhibit large level of errors (in particular $\rho(S_1^C, S_{11}^0)$). These results show that the three different cases provide similar results in terms of accuracy of the calibrated distribution, suggesting that implementing a sequential approach does not downgrade much the estimation of the population parameters. A final comparison can be made to study the computational behavior of the sequential scheme with respect to the method that uses design matrices. Two indicators are studied here: the number of (material) model evaluations and the number of likelihood evaluations, all reported in Figure 4.18 for the three cases over the twenty repetitions.

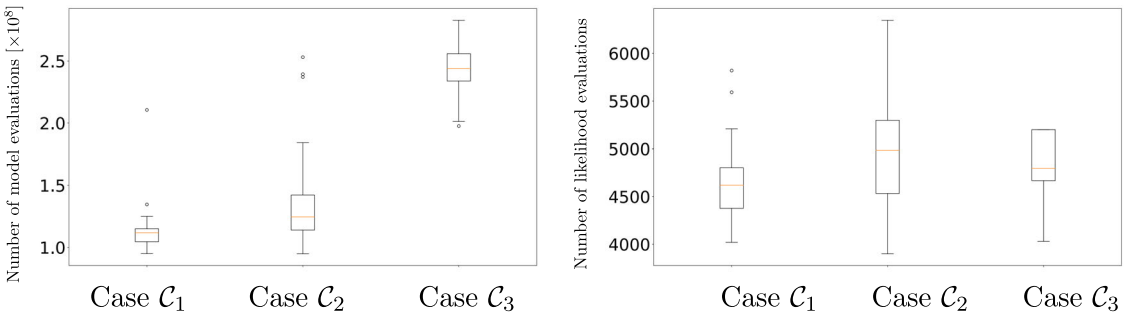
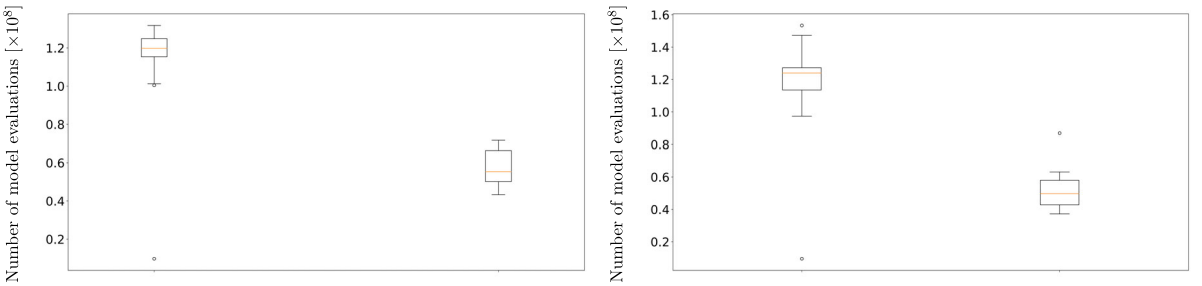


Figure 4.18: On the left, number of model evaluations for the three cases and on the right, number of likelihood evaluations for the three cases. In cases $\mathcal{C}_1$ and $\mathcal{C}_2$, the costs of both stages are combined together.

Figure 4.18 shows that with a sequential scheme, in total over the two phases, the likelihood function is evaluated a comparable number of times in each case, even if in Case $\mathcal{C}_2$ (when ending with the tension experiments), a slight increase can be noticed. Still, the increase remains an within acceptable range. On the contrary, the interest of using a sequential approach arises with the study of number of model evaluations, which is of prime interest. Indeed, the costly part of the likelihood computation refers to the resolution of the minimization problem defined in Eq. (4.11). Focusing on the number of model evaluations demonstrate the interest of using a sequential scheme as over the two phases, the number of model evaluations is approximately divided by two from case $\mathcal{C}_3$ to case $\mathcal{C}_1$ and by 40% from case $\mathcal{C}_3$ to case $\mathcal{C}_2$, which can be considered as a significant gain in terms of computational cost. Still, the number of model evaluations remains high, about 10^8. This number can be partly explained by the choices and settings of the optimization algorithms used for the minimization of both the opposite log-likelihood and function $g_i(\cdot)$. For more complex, it may be necessary to combine the sequential strategy with other approaches such as surrogate modeling to guarantee the appliability of the calibration method. If such solution is chosen, the surrogate model would be easier to train compared to the standard case. Indeed, throughout the sequential process, the search spaces over which population parameters evolve are limited to preserve consistency between the different steps. Therefore, the range of validity of the chosen surrogates can be smaller, making them easier to build, contrary to the standard case in which the entire search space is considered. Note that, as in most procedures implementing surrogate models, it should be checked that the trained surrogate model does not introduce errors that could impair the calibration process. The observed reduction can be explained by different reasons. First, notice in Figure 4.18, for the three cases, the number of likelihood evaluations are (in total for cases $\mathcal{C}_1$ and $\mathcal{C}_2$) similar. In fact, the number of likelihood evaluations in each stage of the sequential approaches is about half of the number of likelihood evaluations when design matrices are used. This can be understood as there more parameters to estimate (17 = 13 population parameters + 4 residual models

with design matrices instead of at most 11 = 9 population parameters + 2 residual models in the application of the sequential strategy with both traction and compression experiments). However, for case $\mathcal{C}_3$, for each likelihood evaluation, the models for both the compression and tension experiments are identified.
On the contrary, for the two other cases, in each stage of the sequential process, the computation of the associated likelihood of mixed-effects requires the estimation of the individual parameters for either the compression or traction experiments but not for both of them at once. This adds to the fact that, in the second step of the methodology, the individual calibrations are eased. Indeed, the mean parameters around which the individual deviations $\boldsymbol{\theta}_i$ are searched (see Eq. (4.11)) are already estimated with low errors (smaller than 1% in average). This is even more important because of the concentration of the linear elastic parameters around their average values. This is illustrated in Figure 4.19, which compares for the twenty repetitions of the calibration process, the number of model evaluations for the second steps of the sequential strategy, *i.e.*, case $\mathcal{C}_1$ with the compression experiments and case $\mathcal{C}_2$ with the tension experiments to their counterpart of case $\mathcal{C}_3$. Indeed, the number of model evaluations is almost divided by two between the case $\mathcal{C}_3$ and the second stages of both cases $\mathcal{C}_1$ and $\mathcal{C}_2$. This significant improvement partly explains the gain observed in terms of model evaluations. A further study on the convergence behavior should be carried out to demonstrate rigorously the interest of the sequential approach compared to classical mixed-effects methods with design matrices.



Compression experiments in case $\mathcal{C}_3$ | Compression experiments in case $\mathcal{C}_1$ | Tension experiments in case $\mathcal{C}_3$ | Tension experiments in case $\mathcal{C}_2$

Figure 4.19: On the left, the total number of model evaluations for the estimation of the individual parameters with the compression experiments in case $\mathcal{C}_3$ and on the second stage of case $\mathcal{C}_1$. On the right, the total number of model evaluations for the estimation of the individual parameters with the tension experiments in case $\mathcal{C}_3$ and on the second stage of case $\mathcal{C}_2$.

This study of the sequential scheme enables to illustrate some advantages and drawbacks. Among the benefits of such a strategy can be cited a proper estimation of the population parameters, at least as accurate as the one that can be obtained with the use of design matrices in this test-case. Furthermore, the number of model evaluations necessary to carry out the calibration of the population parameters is significantly reduced. Finally, it also reduces the number of population parameters of estimated in each calibration problem, allowing to consider more complex phenomena in future works. However, the process exhibits its own limitations. For instance, it could be interesting to test different solutions for the trust regions instead of a simple hypercube. In addition, the routine could have been continued until reaching convergence as in a fixed-point approach at the expense of a larger computational costs. Still, this test-case is interesting as the results can be compared to the standard calibration process that sometimes cannot be carried out because of the computational costs and the dimension of the optimization problem. This method is now applied to experimental data that comes from experiments on a carbon/epoxy material.

4.6 Calibration with real data

In this section, the objective is to apply the sequential scheme implemented before to data that come from real experiments performed on 0°, 90° and ±45° laminates. The data corresponding to the experiments are presented in Section 4.6.1. Numerical settings are briefly discussed in Section 4.6.2. The calibration of the distribution of S_{22} and S_{66} from the responses on the 90° and ±45° laminates, which can be achieved with generalized linear mixed-effects models (see remark in Section 4.4.2) is detailed in Section 4.6.3. Then, in Section 4.6.4, the different possibilities to calibrate the joint distribution of S_{11}^0, S_1^C, S_1^T and ν_{12} are implemented, *i.e.*, the sequential strategy starting with the tension experiments and ending with the compression experiments, the sequential strategy starting with the compression experiments and ending with the tension experiments and finally by considering them all together with the help of design matrices. To illustrate the interest of calibrating such distributions, the estimated variabilities from the different methods are propagated to estimate the failure stress of a perforated plate for several laminates in Section 4.7. The different steps of the analyses are summed up in Figure 4.20.

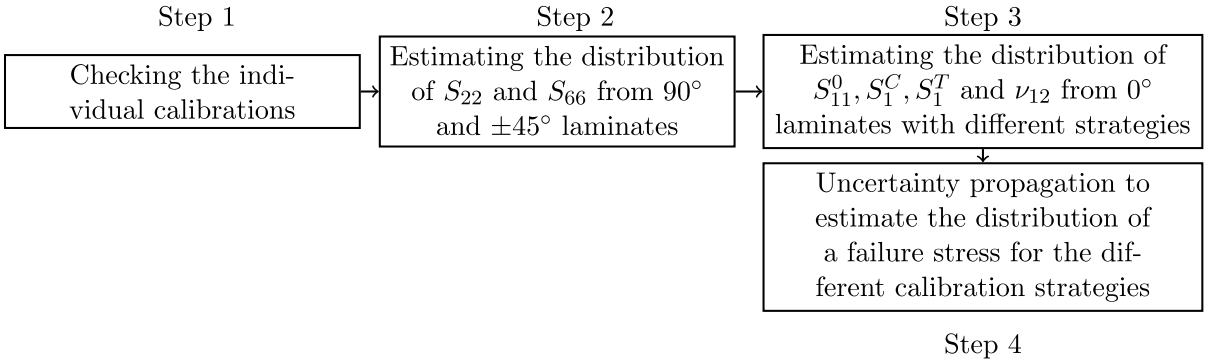


Figure 4.20: Flowchart of the different steps for the assessment of the mixed-effects calibration results.

4.6.1 Data presentation

The database used for the calibration is composed of 32 experiments made on 0° laminates, 16 of them consist in tension tests and 16 of them consist in compression tests. For the identification of the marginal of S_{22}, 15 compression tests made on 90° laminates are processed. The marginal of S_{66} is calibrated with 15 tension tests made on ±45° laminates. For confidential issues, the data depicted here are normalized (between 0 and 1 if it has positive values or between −1 and 0 for negative values). Figure 4.21 depicts the data for the 90° laminates and ±45° laminates.

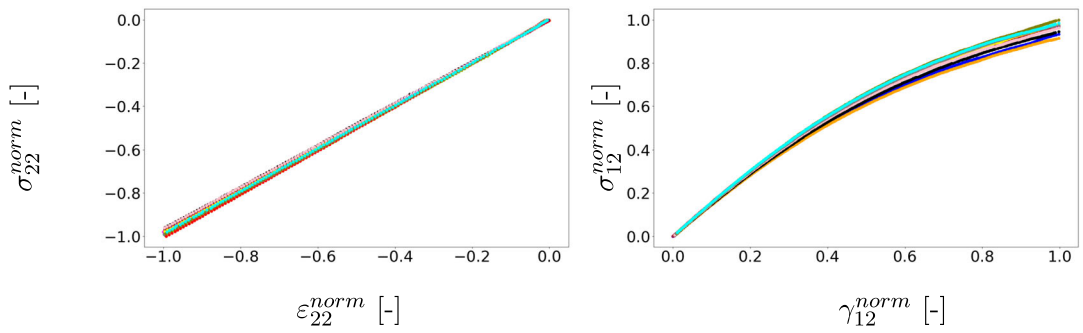


Figure 4.21: Normalized compression tests made on the 90° (on the left) and normalized tension tests made on the ±45° (on the right). The exponent *norm* stands for normalized. Subscripts 1 and 2 refer respectively to the longitudinal and transverse directions of the material axes.

The results of the compression tests made on the 90° exhibit smaller variability compared to the ±45° tests. Still, measurements errors seem sufficiently small to allow material variability to express. The shear strain γ_{12} in the material axes is not directly measurable from the ±45° tests. Yet, it can be recovered as the difference of the longitudinal and transverse strains (in the load axes), which gives $\gamma_{12} := \varepsilon_{11} - \varepsilon_{22}$. Notice that the shear behavior shown by the strain-stress curves is nonlinear, which demands either to adapt the formulation of the elastic linear model (see discussion below) or to cut the experimental curves (that would require a user-defined linearity threshold). The individuals do not have the same number of observation points for both laminates. To make sure that all individuals have the same weight in the likelihood function, all strain-stress curves except the one with the least measurements are periodically sub-sampled while preserving the last point. This procedure is applied for both types of tests. The results from the tension and compression experiments performed on the 0° are presented in Figure 4.22.

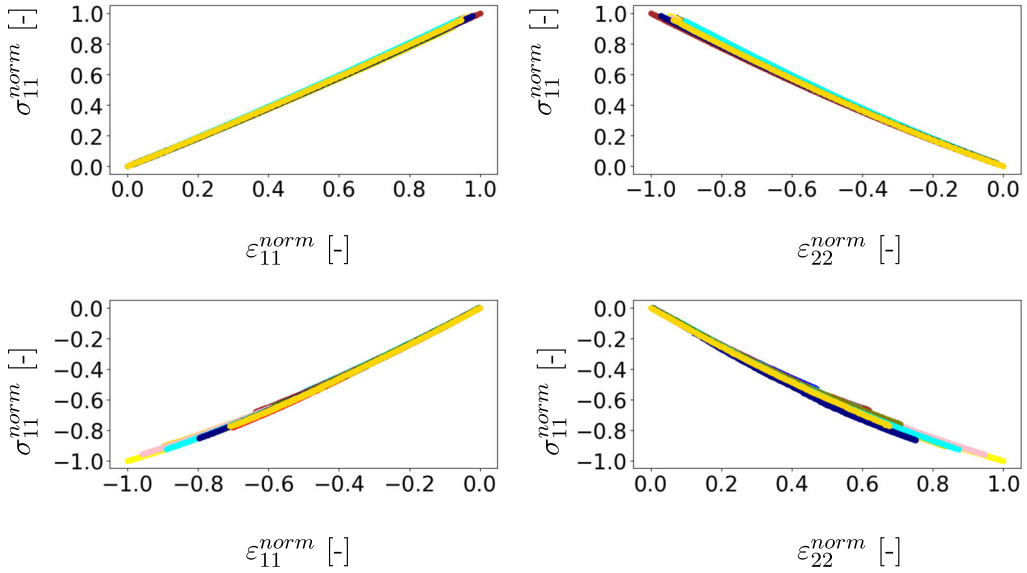


Figure 4.22: Normalized tests made on the 0° laminate with longitudinal (first column) and transverse outputs (second column). On the first row, tension results and on the second row, compression results. The exponent *norm* stands for normalized. Subscripts 1 and 2 refer to the longitudinal and transverse directions of the material axes.

The results of the tension tests made on the 0° exhibit smaller variability than the compression experiments. Still, measurement errors are negligible in tension but appear to be larger in compression. This is not surprising as compression experiments are harder to conduct and other phenomena such as buckling can appear during the experiment. As for the previous experiments, it is necessary to sub-sample some of the tests to make sure that all of them will have similar influence in the likelihood function.

4.6.2 Numerical settings

The settings for the minimization of the log-likelihood remains the same as those described in Section 4.6.2. Still, experimental data present inherent difficulties that make the individual calibrations harder, in particular the presence of model inadequacy. Finally, to ensure the consistency of the results, the optimization is repeated 4 times with different initialization points, namely $\gamma \times \mathbf{1}_{\mathbb{R}^p}$ with $\gamma \in \{0.15, 0.40, 0.65, 0.90\}$. The number p depends on the joint distributions to calibrate, that are detailed in the appropriate sections. Before proceeding to the calibration, it is possible to check to what extent the individual calibration succeed for the 0° laminates (the behavior of 90° and ±45° experiments is described by univariate models for which the calibration of the individual parameters does not present issues). As there is no exact value for the individual

parameters, we can compute a RMSE between the experimental and calibrated longitudinal and transverse strains. For 1,000 $\mathbf{\Psi}$ uniformly drawn within the bounds defined in Table 4.3, the individual parameters for all experiments are estimated and depicted in Figure 4.22.

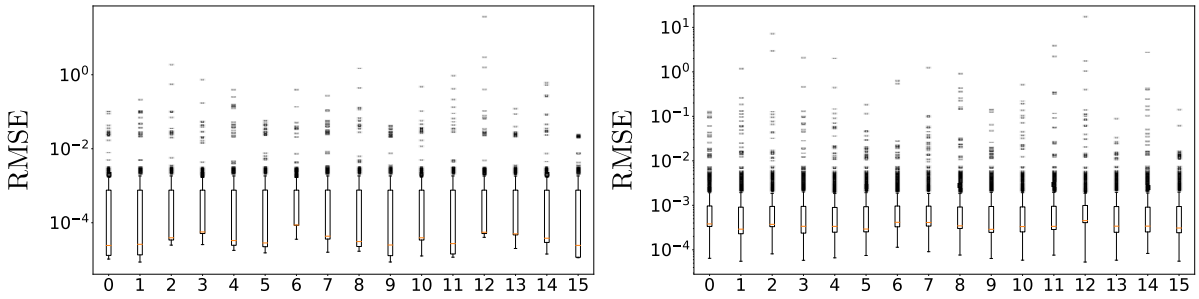


Figure 4.23: For 1,000 candidates $\mathbf{\Psi}$, normalized RMSEs between the experimental and calibrated curves in traction (on the left) and in compression (on the right) for all UD 0° experiments (they are indicated on the x-axis).

Figure 4.23 shows that for all experiments, for most candidate of full model parameters, the individual minimizations provide accurate estimations of the experimental curves. Indeed, most RMSEs remain below 10^{-2} which is rather small compared to the level of the output curves that lies here between 0 and 1 as depicted in Figure 4.22. The sequential calibration process can be now implemented to the available experimental data.

4.6.3 Calibration of the marginals of the transverse and shear elasticity

This section presents the calibration of the marginals of S_{22} and S_{66}. To infer the marginal of S_{22}, the model employed to mimic the strain-stress curves of 90° laminates is simply $\varepsilon_{22} = S_{22}\sigma_{22}$, with S_{22} the linear compliance whose marginal has to be estimated. The modeling of the shear strain ε_{12} is slightly more complicated. Indeed, as illustrated in Figure 4.21, the strain-stress relation is nonlinear, that can be modeled with complicated models involving the resolution of differential equations as in more advanced models such as the ONERA Progressive Failure Model [Laurin, 2005]. However, as here only the linear part should be properly estimated, a simpler approach consists in approximating the strain-stress curve with a linear behavior and a nonlinear function, for instance, taken as the power function which gives the following relation [Hahn and Tsai, 1973]:

$$\gamma_{12} \coloneqq \frac{S_{66}}{2}\sigma_{12} + \alpha\sigma_{12}^{r} \ , \tag{4.27}$$

with r the power of the stress (a shape parameter, without unit and specific physical meaning) and α, a normalizing coefficient in MPa^{-r} (with no specific physical meaning). The exponent can be either calibrated ([Swanson et al., 1985]) or fixed to a nominal value (3 for instance by Camanho [Camanho et al., 2006]). In this case, it is decided to set it to 5, a value that provide accurate fit of the strain-stress curves following expert recommendation. To limit the number of parameters to estimate with mixed-effects, a single residual model for all experiments with both laminates is chosen. This makes 3 parameters to estimate, all assembled within $\mathbf{\Psi}_{90°}$, the set of parameters inferred from the 90° laminate. They stand for $\mathbf{\Psi}_{90°} = [\mu_{S_{22}}, \mathrm{sd}(S_{22}), \omega_{90°}]$, with $\omega_{90°}$ the variance of the residual model that is identical for all experiments. With the ±45° laminates, there are 6 parameters to estimate, all assembled within $\mathbf{\Psi}_{\pm 45°}$, namely $\mu_{S_{66}}, \mathrm{sd}(S_{66}), \mu_{\alpha}, \mathrm{sd}(\alpha), \rho(S_{66}, \alpha)$ and $\omega_{\pm 45°}$. The estimated marginals with the two laminates are reported in Table 4.16.

Table 4.16: Calibrated marginal parameters $\widehat{\mathbf{\Pi}}_{90^\circ}$ from the 90° laminate and calibrated marginal parameters $\widehat{\mathbf{\Pi}}_{\pm 45^\circ}$ from the ±45°. The coefficients of variation of the calibrated population parameters, lower than 1%, are omitted for the sake of clarity.

	S_{22} [$\times 10^{-6}$MPa^{-1}]	S_{66} [$\times 10^{-6}$MPa^{-1}]	α [$\times 10^{-13}$MPa^{-5}]
Calibrated means	116	281	7.80
Calibrated standard-deviations	1.43	6.14	0.848

The calibrated correlation between S_{66} and α stands for 0.859. This value is difficult to validate/refute from a physical point of view as this coefficient does not have any particular meaning, its only role is to catch the nonlinearity that comes from the viscous behavior. On the contrary, it is possible to confront the values of the compliances to references available in the literature for materials close to the one studied here. The values in Table 4.16 appear to be consistent with other values available in the literature for similar materials. For instance, in Rollet [Rollet, 2007], the average of S_{22} corresponds to 1.19×10^{-4} and 1.25×10^{-4} in Laurin [Laurin, 2005]. To check the calibration, it is necessary to study the individual fit. This can be done by plotting for one repetition of each type of test, the experimental and calibrated strain-stress curves, as in Figure 4.24. The population parameters calibrated for the 4 repetitions are almost the same. Then, provided that the individual parameters are determined by resolving almost always the same minimization problem, as one could expect, the individual parameters estimated for each repetition of the calibration process exhibit negligible differences. Thus, for the sake of clarity, only the individual parameters corresponding to the best in terms of likelihood are presented.

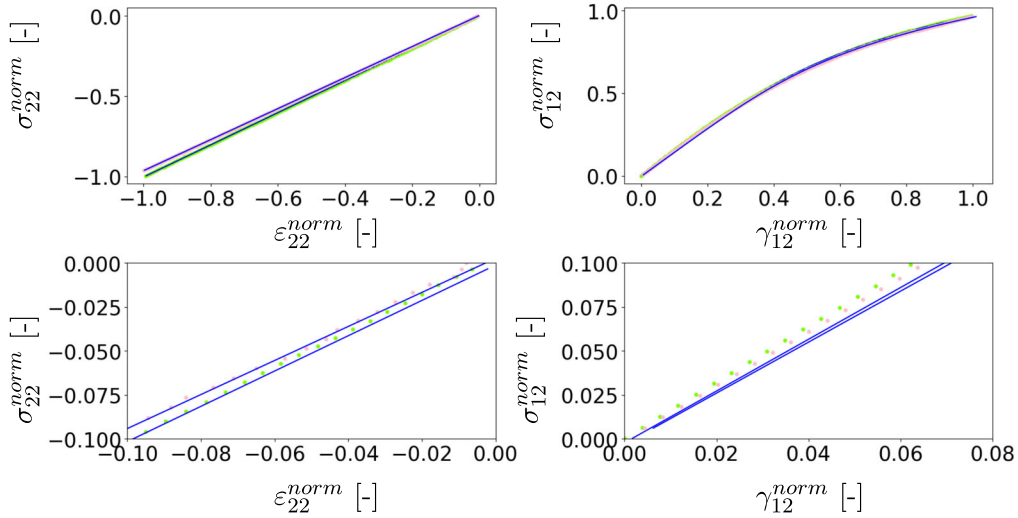


Figure 4.24: For two repetitions randomly selected (the 6$^{\text{th}}$ and the 13$^{\text{th}}$), normalized experimental (dots) and calibrated responses (blue lines) for the compression tests made on the 90° (left column) and normalized tension test made on the ±45° (right column). The second row focuses on the beginning of the strain-stress curves. The exponent *norm* stands for normalized. Provided their small variabilities, only the models corresponding to the best calibration in terms of likelihood are presented.

In Figure 4.24, the calibration of the experiments for the response on the 90° laminates provides satisfying results. The calibration of the ±45° laminates does not achieve the same level of accuracy. In particular, a slight under-estimation of the linear compliance at the beginning of the strain-stress curve (see the bottom right picture of Figure 4.24), along with an overestimation of the strain at the end of the experiment for the ±45° laminates appear, but still, the results provided remain acceptable. This can be due to the fact that a very simple model is

used to mimic a complex phenomenon. Numerically, with the normalized values, the normalized calibrated variance of the residual model ω_{90° reaches 1.25×10^{-3} (to be compared to 1) and $\omega_{\pm 45^\circ}$ corresponds to 5.40×10^{-3} (also to be compared to 1), all of them with a coefficient of variation lower than 1%. These values are in agreement with the estimated noises that have similar values, assessing the accurate estimation of the available experiments, confirmed by the raw residuals $\widehat{\mathbf{r}}_i := \mathbf{y}_i - \mathcal{F}(\mathbf{t}, \widehat{\boldsymbol{\theta}}_i)$ depicted next for the calibrations on both the 90° and ±45° experiments.

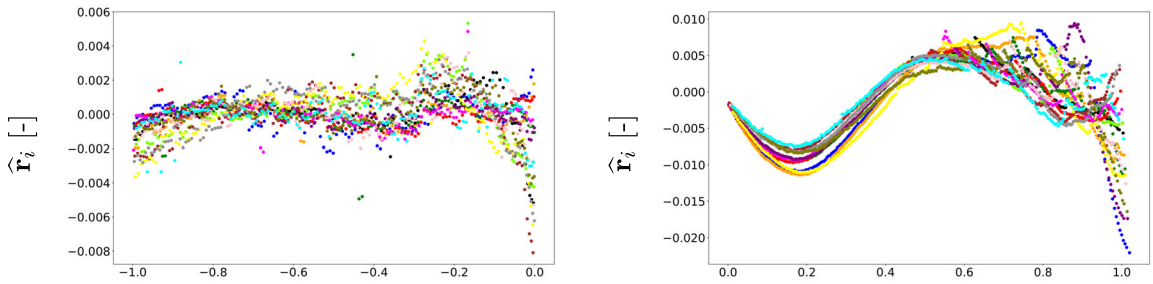


Figure 4.25: For all repetitions, normalized raw residuals between the normalized experimental and calibrated data from 90° (on the left) and ±45° tests (on the right).

Figure 4.25 points out the presence of significant model inadequacy that comes from different sources according to the type of experiments. For the 90°, it is significant at the beginning of the curve and quickly becomes much smaller for the rest of the experiment. In particular, a remaining structure can be spotted but its magnitude is comparable if not smaller than oscillations that embody measurements errors. The presence of bias at the start of the experiment can be partly explained by the fact that at low stress, the axis of the specimen is not perfectly mixed up with the loading axis of the testing device. Otherwise, the range of the residuals seems constant throughout the experiments, suggesting that the homoscedastic assumption can be validated to a certain extent. On the contrary, for the ±45° laminates, the residuals exhibit a clear structure that demonstrates the presence of the model discrepancy. This bias comes from the fact that a simple model defined in Eq. (4.27) is not able to catch all features of experimental curves.
In Figure 4.25, in average, the residuals represent at most 2% of the model output, hinting for a proper estimation of the average behavior. However, with mixed-effects, the objective is not only to reproduce a global trend but also the dispersions around it. Because of the high level of model inadequacy, we may wonder whether in this case, model bias is really negligible compared to material variability. In other words, are the population approaches a relevant frame to handle the calibration of the repetitions of the ±45° experiments ? To answer this question, we should check whether material variability is necessary to explain the discrepancies between the model curves or if the combination of measurements errors and model bias is sufficient. To investigate this point, it is possible to reproduce the same analysis that was implemented in Chapter 1 to highlight the presence of material variability. Recall that the first step consists in quantifying independently on each specimen the impact of measurements errors on the individual parameters, which can be done by sampling the distribution $f(\boldsymbol{\theta}_i|\mathbf{y}_i, \widehat{\boldsymbol{\Pi}}, \widehat{\omega}_i)$, whose PDF is proportional to $f(\mathbf{y}_i|\boldsymbol{\theta}_i, \widehat{\boldsymbol{\Pi}}, \widehat{\omega}_i) f_{\boldsymbol{\Theta}|\widehat{\boldsymbol{\Pi}}}(\boldsymbol{\theta}_i|\widehat{\boldsymbol{\Pi}})$, allowing the use of MCMC algorithms among which can be found the Metropolis-Hastings algortihm [Metropolis et al., 1953, Hastings, 1970] implemented for instance in the emcee module [Foreman-Mackey et al., 2013] (the measure on 90° and ±45° tests are univariate, hence the use of $\mathbf{y}_i$ and ω_i instead of $\widetilde{\mathbf{Y}}_i$ and $\widetilde{\boldsymbol{\Omega}}_i$). The thousand first samples are burned-in and one every 30 samples available is kept (thinning). The samples are denoted $\tilde{\boldsymbol{\theta}}_i$ and their total number is labeled N_{MC}, which gives the following sets of invidual parameters $\{\tilde{\boldsymbol{\theta}}_{\ell,i}^{\text{MCMC}}\}_{\ell \in [\![1, N_{MC}]\!]}$. This step is carried out with two specimens (the sixth and the thirteenth for each type of experiment). In a second stage, the bundle of model curves for the parameters sampled with MCMC techniques are compared. However, in this particular case, because the

curves are stucked together, it is more relevant to compare the differencies between the data for one experiments and the bundle of model curves for the MCMC samples determined in the previous steps. To be more precise, the trajectories corresponding to the difference between the experimental data for one specimen (for instance the 6^{th}) and the model for the MCMC samples $\{\tilde{\boldsymbol{\theta}}_{\ell,6}^{\text{MCMC}}\}_{\ell\in[\![1,N_{MC}]\!]}$ determined on the 6^{th} experiment, that is to say $\tilde{\mathbf{r}}_6 = \mathbf{y}_6 - \mathcal{F}(\mathbf{t}_6, \tilde{\boldsymbol{\theta}}_{\ell,6}^{\text{MCMC}})$, are computed. These trajectories are then compared to those of the difference between the experimental data for the 6^{th} specimen and the model for the MCMC samples $\{\tilde{\boldsymbol{\theta}}_{\ell,13}^{\text{MCMC}}\}_{\ell\in[\![1,N_{MC}]\!]}$ determined on the 13^{th} experiment, that is to say $\tilde{\mathbf{r}}_{6,13} = \mathbf{y}_6 - \mathcal{F}(\mathbf{t}_6, \tilde{\boldsymbol{\theta}}_{\ell,13}^{\text{MCMC}})$. Those one are depicted in Figure 4.26 for both the UD 90° and ±45° laminates. In Figure 4.26, $N_{MC} = 1{,}000$.

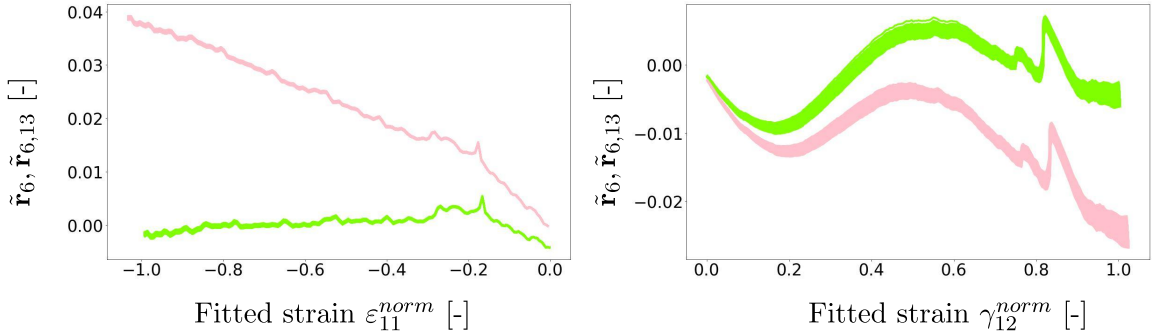


Figure 4.26: For all repetitions, bundle of normalized raw residuals (in light green) between the normalized experimental and calibrated data from 90° (on the left) and ±45° laminates (on the right) compared to the trajectories of the normalized differences between the experimental data of the 6^{th} specimen and the model for the MCMC samples determined on the 13^{th} experiments (in pink).

Figure 4.26 shows that for both the 90° and the ±45° tests, there is a clear gap between the bundle of trajectories of the normalized raw residuals and the bundle of normalized trajectories between the experimental data of the 6^{th} specimen and the model for the MCMC samples determined on the 13^{th} experiment. It shows that neither model bias nor measurements errors can explain the gap between the different repetitions, which enables the intrinsic variability between the specimens to express. This asserts the presence of material variability and justifies the use of the population approaches. In fact, even if the assumptions on the residuals are not met, it does not prevent a proper estimation of the individuals (either for the trend or the dispersions around) and consequently, of the model parameters distribution.

4.6.4 Calibration of the joint distribution of the longitudinal elasticity

This section presents the results about the calibration of the joint distribution of S_{11}^0, S_1^C, S_1^T and ν_{12} either implementing the sequential strategy or using design matrices. First, a focus on the results provided with the sequential strategy is made as with the synthetic data. In the first phase, with the tension experiments, the optimization variables are $\mu_{S_1^T}, \mu_{S_{11}^0}, \mu_{\nu_{12}}, \mathbb{V}(S_1^T), \mathbb{V}(S_{11}^0), \mathbb{V}(\nu_{12})$ and optionally, $\rho(\nu_{12}, S_{11}^0), \rho(\nu_{12}, S_1^T), \rho(S_1^T, S_{11}^0)$, all assembled within $\mathbf{\Pi}_{UD0^\circ}^T$ (the exponent T stands for tension). A single residual model is chosen for all the experiments. Though it is not the most appropriate assumption, considering different residual models for each of the experiment would bring much complexity. Indeed, as there are sixteen individuals with 2 output measures each, this would require the estimation of $6 + 32 = 38$ parameters (without a dependence structure) and $9 + 32 = 41$ parameters (with a dependence structure). This makes the optimization complex to perform. Consequently, to limit the number of parameters to calibrate, it is decided to keep the assumption of a single residual model for all experiments. The estimates of the population parameters with the tension experiments are reported in Tables 4.17 and 4.18, considering either a dependence structure or not.

Table 4.17: Calibrated marginal parameters $\widehat{\mathbf{\Pi}}^T_{UD0^\circ}$ from the 0° laminate with the tension experiments, with and without dependence structures. The coefficients of variation of the calibrated population parameters, lower than 1%, are omitted for the sake of clarity.

	S^0_{11} [$\times 10^{-7}$MPa^{-1}]	S^T_1 [$\times 10^{-7}$MPa^{-1}]	ν_{12} [$\times 10^{-2}$]
Calibrated means without dependence structure	87.3	42.7	30.7
Calibrated means with dependence structure	87.3	42.7	30.7
Calibrated standard-deviations without dependence structure	0.980	0.835	0.316
Calibrated standard-deviations with dependence structure	0.978	0.834	0.316

Table 4.18: Calibrated correlation parameters $\widehat{\mathbf{\Pi}}^T_{UD0^\circ}$ from the UD 0° laminate with the tension experiments. The coefficients of variation of the calibrated correlations, lower than 1%, are omitted for the sake of clarity.

	$\rho(\nu_{12}, S^0_{11})$	$\rho(\nu_{12}, S^T_1)$	$\rho(S^T_1, S^0_{11})$
Calibrated correlations	0.323	0.242	0.323

The averages of the elastic linear coefficients (*i.e.*, S^0_{11} and ν_{12}) are consistent with typical values for this parameter, as for instance those proposed by [Laurin, 2005]. It is harder to interpret the calibrated values of the S^T_1 provided that it is a shape parameter that depends on the choice of ε^T_0. Its average is lower than the one of S^0_{11}, which is expected in theory. The correlations are difficult to analyze from a physical point of view. Indeed, two of them are weak ($\rho(\nu_{12}, S^0_{11})$ and $\rho(\nu_{12}, S^T_1)$) and it is difficult to decide whether they do represent a physical phenomenon or are simply used to improve the likelihood function without proper meaning, considering the small set of individuals available. Remark that adding a dependency structure does not change the values of the standard-deviation. Remember also that both the correlations and standard-deviations express in the covariance matrix $\mathbf{\Sigma}$, *i.e.*, in the regularization part of function $g_i(\cdot)$ defined in Eq. (4.9) and coerce the individual parameters to be consistent with $f_{\mathbf{\Theta}|\mathbf{\Pi}}$. Thus, if adding extra constraints on the individual parameters does not clearly change the estimated standard-deviations, the inclusion of a dependency structure may be questioned. Small differences between the standard-deviations of S^0_{11} and S^T_1, associated to the highest correlation, can still be observed. This could suggest that this correlation is significantly different from 0. Yet, provided the small number of individuals and the approximations, this analysis should be taken cautiously. In fact, the use of correlation do not seem justified by the data. The Bayesian Information Criterion (BIC) which can be used to compare different models for $f_{\mathbf{\Theta}|\mathbf{\Pi}}$ equals $-39{,}101$ without dependency structure and $-39{,}106$ with correlations included. Generally, the modeling to which the lowest BIC is associate stands for the best one. However, this criterion is asymptotic (with the number of individuals). Thus, given the small difference between (lower than 0.1%) and the number of available individuals, the two models are in fact equivalently likely and there is no specific justification in the modeling of dependencies. Thus, in the following, we will only discuss the calibration without correlations.

In Figure 4.27, the calibrated models appear to fit accurately the strain-stress curves the longitudinal responses. On the transverse responses, the calibration does not reach the same level of accuracy. This is not surprising: the transverse response is mostly influenced by the matrix whose behavior comprises a viscous component not taken into account here (partly because of computational costs reasons). Numerically, the normalized calibrated standard-deviation of the noise for the longitudinal experiments is about 1.22×10^{-3} and reaches 1.11×10^{-2} (to be com-

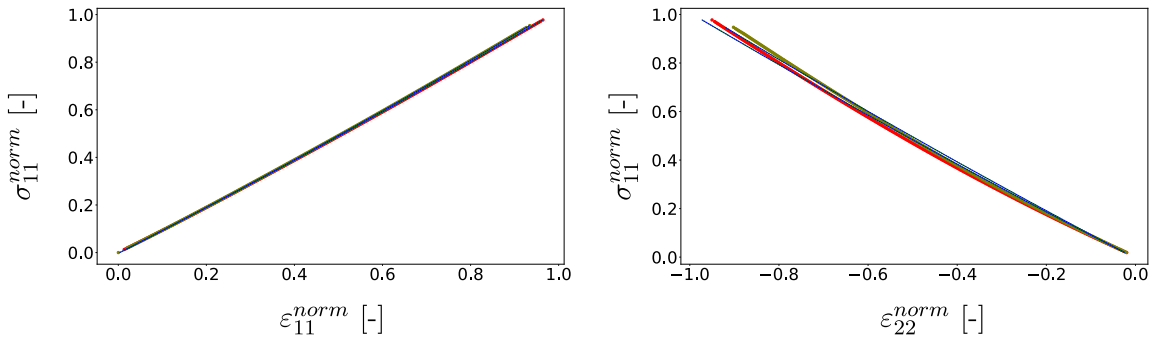


Figure 4.27: For two test repetitions (the 6$^{\text{th}}$ and the 13$^{\text{th}}$), normalized tension experimental data from 0° laminates (dots) and corresponding calibrated model (black lines for the individual parameters identified without dependence structure and red dashed lines for the model identified with a dependence structure). On the left, the longitudinal responses and on the right, the transverse responses.

pared to 1). These values are in agreement with the estimated noises that have similar values, assessing the accurate estimation of the available experiments. It also appears that the models for the individual parameters estimated with and without dependency structures are almost the same, hinting that the individuals are very similar (the difference between them is smaller than 10^{-2}%). As earlier, the raw residuals $\widehat{\mathbf{r}}_i := \mathbf{y}_i - \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)$ are depicted in Figure 4.28 for the calibration of the longitudinal and transverse outputs on the 0° laminate.

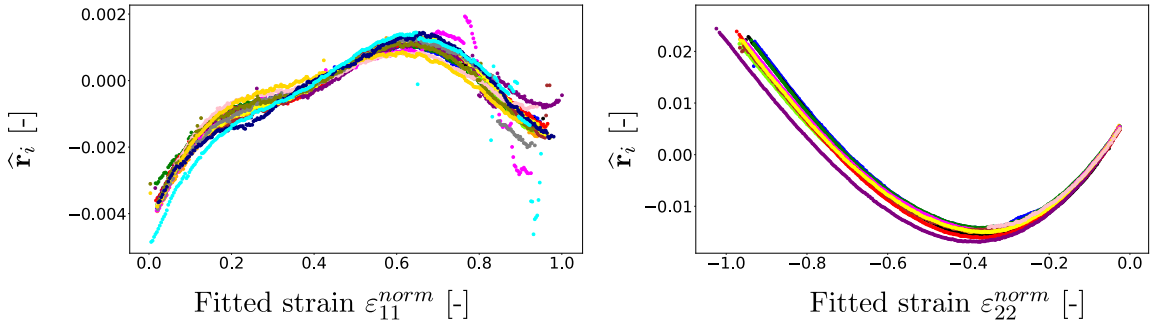


Figure 4.28: For all repetitions, normalized raw residuals between the normalized experimental and calibrated data from 0° laminates for the longitudinal (on the left) and transverse strains (on the right).

The raw residuals confirm the accurate calibration of the experiments as the residuals represent in average only 2% of the model outputs, pointing out that the global trend is accurately estimated. However, again a structure within the residuals can be spotted, highlighting the presence of significant model bias, in particular for the transverse response. It is therefore to wonder again whether model bias does not outweight the impact of the intrinsic variability of the available specimens. To investigate this point, the process that was implemented in Section 4.6.3 can be repeated. The obtained bundle of curves are depicted in Figure 4.29.
As earlier, Figure 4.29 shows that for both the longitudinal and transverse responses on UD 0°, a clear gap between the bundle of trajectories of the normalized raw residuals and the bundle of normalized trajectories between the experimental data of the 6$^{\text{th}}$ specimen and the model for the MCMC samples determined on the 13$^{\text{th}}$ experiment can be spotted. It shows that neither model bias nor measurements errors can explain the gap between the different repetitions, allowing the intrinsic variability between the specimens to express. This asserts the presence of material variability and justifies the use of the population approaches. In fact, even if the assumptions on the residuals are not met, it does not prevent a proper estimation of the individuals (either for the trend or the dispersions around) and consequently, of the model parameters distribution. Yet, an obvious potential improvement would consist in accounting for this model inadequacy.

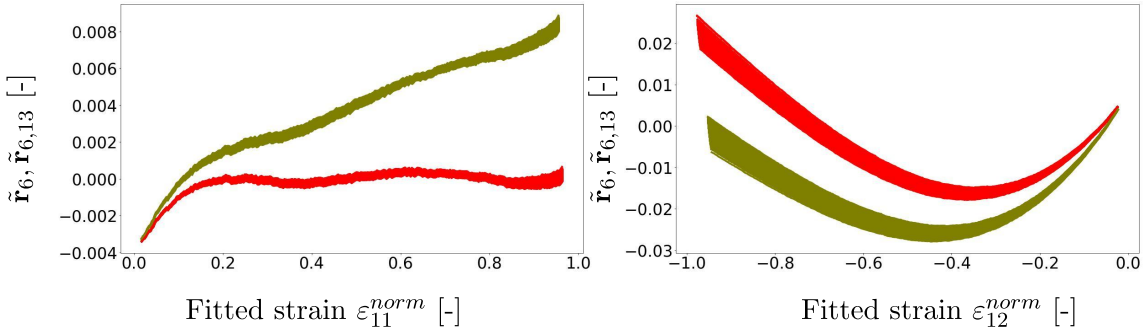


Figure 4.29: For all repetitions, bundle of normalized raw residuals (in red) between the normalized longitudinal experimental and calibrated data (on the left) and transverse (on the right) responses from 0° tests compared to the trajectories of the normalized differences between the experimental data of the 6$^{\text{th}}$ specimen and the model for the MCMC samples determined on the 13$^{\text{th}}$ experiments (in dark green). 1,000 MCMC samples are propagated.

The second step of the methodology can be now carried out. In this phase, the objective is to estimate the population parameters that monitor the compression behavior, that is to say $(\mu_{S_1^C}, \mathbb{V}(S_1^C))$ and to update those that characterize the elastic linear parameters ($\mu_{S_{11}^0}, \mu_{\nu_{12}}, \mathbb{V}(S_{11}^0)$, $\mathbb{V}(\nu_{12})$). These parameters are limited first to a search space that is $\mathcal{R}_1 = [0.8\widehat{\mathbf{\Pi}}_{0^\circ}^T, 1.2\widehat{\mathbf{\Pi}}_{0^\circ}^T]$. The same residual model as before (a single residual model for all experiments) is chosen to limit the number of parameters to calibrate. Without dependence structure, 8 parameters should be estimated and with a dependence structure included, 12 parameters have to be estimated. The estimates of the population parameters, $\mathbf{\Pi}_{0^\circ}^C$, with the compression experiments are reported in Tables 4.19 (all parameters are assumed as independent).

Table 4.19: Calibrated marginal parameters $\widehat{\mathbf{\Pi}}_{0^\circ}^C$ from the 0° laminate with the tension experiments without dependence structures. The coefficients of variation of the calibrated population parameters, lower than 1%, are omitted for the sake of clarity.

	S_{11}^0 [$\times 10^{-7}$MPa^{-1}]	S_1^C [$\times 10^{-7}$MPa^{-1}]	ν_{12} [$\times 10^{-2}$]
Calibrated means	87.6	498	35.2
Calibrated standard-deviations	2.79	42.8	0.874

The calibrated marginals of the S_1^C is complicated to assess, except noticing that the mean of the S_1^C is greater than S_{11}^0, which is consistent from a physical point of view. On the correlations, the same comments as those with the tension experiments can be made. Let's now concentrate on the joint distribution of S_{11}^0 and ν_{12}. Tables 4.17 and 4.19 show that the averages of both S_{11}^0 and ν_{12} have changed. Still, both of them belong to the search space $\mathcal{R}_1$. If there is almost no change for the mean of S_{11}^0 (a difference smaller than 1%), there is a significant change for the mean of ν_{12} (a difference of 10%) and even clearer for the standard-deviations of these two parameters, for which a change over 100% is observed (note that they are outside trust region $\mathcal{R}_1$). Two different explanations can be proposed. First, remember that the estimation of the population parameters is deterministic. As with few individuals, the uncertainty on the population parameters is significant (as shown in the previous chapter), ignoring such uncertainty can lead us to say that the calibrated population parameters are different in tension and compression. Still, if such uncertainty quantification step was carried out, it could arise that the confidence intervals on the population parameters in tension and compression are not disjoints. We do not carry out such analysis for computational costs reasons. The second reason that could explain the differences between the calibration results to a certain extent is the presence of a significant model bias. In fact, to draw clear conclusions, it would be necessary to set up an appropriate residual model which could be validated contrary to the one used so far. Similar comments can

be made on the individual calibrations about the estimation of the average behavior, about the presence of model bias and about its magnitude with respect to material variability.
These two phases provide population parameters estimates that, despite of the uncertainty on their values, remain different, especially for the standard-deviations. Furthermore, both of them suffer from the same weaknesses (significant presence of model inadequacy) and there is no specific reason to prefer the estimations from either the compression or the tension experiments. Furthermore, if the calibration of this sequential strategy is carried out in the reverse order (starting with compression experiments and then ending with the tension), the results are the same but in the reverse order. To be more precise, the marginals are the same as those reported in Tables 4.19 for the compression phase, and for the tension phase, they are the same as those in Tables 4.17. In fact, there seem to be a global minima for the negative log-likelihood that attracts the population parameters to the same place all the time. Thus, here, the estimation of the population parameters heavily depends on the order of the calibration, which underlines the necessity to test the different possible orders to be able to draw reliable conclusions. To lower this effect, one possibility would be to implement a fixed-point approach that could be considered in future works.

The estimation of the joint distribution of S_{11}^0, S_1^C, S_1^T and ν_{12} can be done either by implementing the sequential strategy (which corresponds to the above results) or with the help of design matrices. With this second method, the population parameters, that refer to the means, standard-deviations and correlations reported in Eq. (4.17) can be calibrated straightforwardly. This makes 13 population parameters with a dependence structure and 8 without (remember that the correlation between S_1^T and S_1^C cannot be estimated and is set to 0). Furthermore, provided the small improvement difference between the BICs with and without dependency structure combined with the small number of individuals, the analysis is here limited to the case where no correlations are considered. It remains to choose the residual model. First, it is chosen to differentiate the tension and compression experiments with different models. To limit the number the parameters, it is decided to consider a single residual for all individuals, which requires to calibrate 4 parameters. Finally, with the chosen modeling, this gives 12 parameters to estimate. If different residual models for each individuals are considered, this is synonymous to 64 standard-deviations parameters to estimate (32 experiments with two output responses each), so in total, 72 parameters, which is almost intractable. The calibrated population parameters $\widehat{\mathbf{\Pi}}_{UD0^\circ,\text{full}}$ are reported in Table 4.20.

Table 4.20: Calibrated marginal parameters $\widehat{\mathbf{\Pi}}_{UD0^\circ,\text{full}}$ from the UD 0° laminate with the tension and compression experiments, without dependence structure. The coefficients of variation of the calibrated population parameters, lower than 1%, are omitted for the sake of clarity.

	S_{11}^0 [$\times 10^{-7}$MPa^{-1}]	S_1^T [$\times 10^{-7}$MPa^{-1}]	S_1^C [$\times 10^{-7}$MPa^{-1}]	ν_{12} [$\times 10^{-2}$]
Calibrated means	87.4	42.6	498	33.0
Calibrated standard-deviations	2.06	0.978	42.7	2.35

In Table 4.20, the marginals of both S_1^T and S_1^C correspond more or less to those calibrated independently in each phase (tension and compression), see Tables 4.17 and 4.19 (the differences are lower than 1%). The small differences come from the small number of experiments resulting in high levels of uncertainty. This is not the case for the linear elastic parameters, S_{11}^0 and ν_{12}. For instance the mean of the Poisson's ratio increases to 0.33, which corresponds to the average of the mean values estimated in the tension and compression phases. In fact, it appears that, such as with the synthetic data, the calibration makes a compromise between the two types of experiments. For S_{11}^0, the change is limited as both type of experiments provided similar estimates of the mean parameter. More precisely, it has some regularizing effect. Indeed, with the

compression experiments, the standard-deviation of S_{11}^0 reached 2.80×10^{-7} against 9.71×10^{-8} with the tension experiments. Then, combining both types of experiments reduces this value, as it occurs when observations with small dispersions are added to others with a comparable average that exhibit greater variability. However, if data originates from two sets of observations with different averages and small standard-deviations in both sets, the opposite phenomenon can be observed as can be noticed with the Poisson's ratio. Indeed, though the averages are close, provided the concentration of the Poisson's ratio around the mean values (the coefficients of variation are about 1% in tension and 2% in compression), the combination of the tension and compression experiments leads to a significant increase of the standard-deviation when both types of experiments are considered, as illustrated in Figure 4.30.

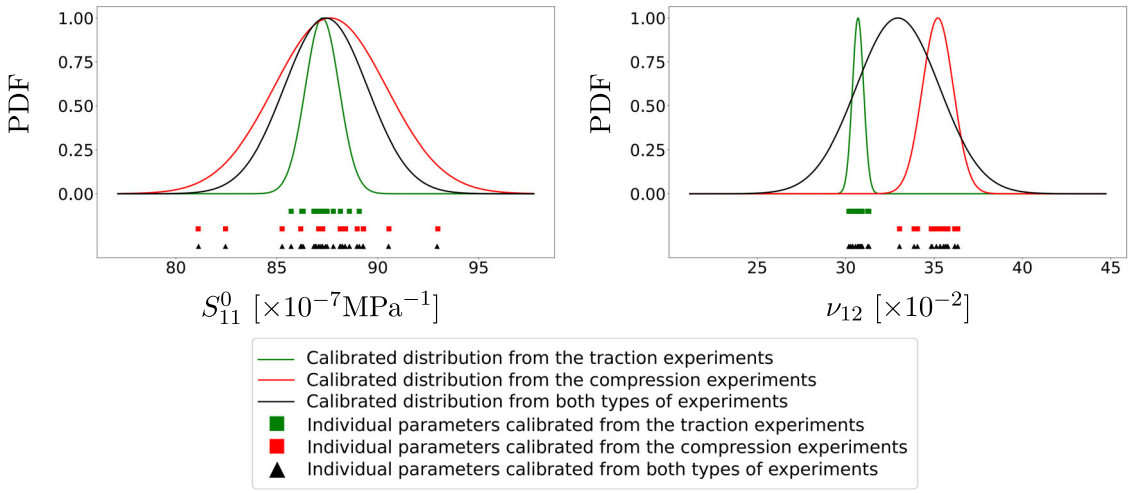


Figure 4.30: PDF of S_{11}^0 and ν_{12} and the corresponding individual parameters for the calibration in the longitudinal direction. For the calibration that used the tension or compression experiments only, the PDF and the individual parameters refer to those calibrated without dependence structure as they are almost identical to those calibrated with dependency structure.

On the right of Figure 4.30, it is worth noticing that the support of the distributions of the Poisson's ratio is different between the tension and compression experiments. Aside form the impact of uncertainties on the population parameters, as stated earlier, it may suggest that this parameter may be different for both types of experiments, hinting for an update of the model formulation. Such a conclusion questions to some extent the accuracy of the estimation of material variability based on a model whose formulation is challenged. On the left side of Figure 4.30, the parameter values seem consistent between the tension and compression experiments but those from the latter exhibit larger variability. This extra dispersion may stem from the higher difficulty to perform compression experiments because of some undesirable phenomena such as buckling. To finish the analysis, the computational costs can be compared between the three methods as summarizes Table 4.21.

Table 4.21: For all repetitions of the calibration process and strategies, numerical costs of the employed method.

Calibration scenario	Total number of model evaluations [$\times 10^7$]	Total number of likelihood evaluations [$\times 10^3$]
Case $\mathcal{C}_1$	$[2.57, 2.71]$	$[3.59, 3.74]$
Case $\mathcal{C}_2$	$[3.29, 3.62]$	$[3.40, 4.06]$
Case $\mathcal{C}_3$	$[4.89, 6.01]$	$[2.96, 3.59]$

From Table 4.21, the same conclusion drawn from Table 4.18 arises, that is to say, with the help of the sequential strategy, it is possible to estimate with much fewer model evaluations (decreased by at least 33%). From a computational point of view, the sequential strategy is the most appealing method material variability is characterized with much fewer model runs than when using design matrices. However, the gain in terms of computational time should not be the only criterion to select the "best" method and we should also rely on a physical analysis to make our choice. Aside from the fact that there is no available database that can be used to confront in particular the standard-deviation values, it should be remembered that because of the lack of specimens, it is highly likely that the population parameters are flawed with uncertainties and comparing point estimates is in fact not relevant. Furthermore, note that the calibration strategy that uses the largest amount of data for linear elastic coefficients does not necessary strengthen their estimations. Indeed, when using design matrices, the calibration process makes trade-offs between the different types of experiments, which can impair the overall identification of the corresponding population parameters if the estimation of the associated individual parameters is difficult to carry out on one type of experiments (for instance because flawed with significant uncertainties) and easier on the other tests. In fact, a more appropriate way to study the impact of the calibration strategy is to propagate the estimated distributions in order to estimate sizing quantity, for instance a failure criterion. This topic is developed in the next section.

4.7 Propagating the distributions

The distributions of the properties of the base ply are not of great interest if considered alone. Indeed, the composite applications almost exclusively use structures consisting of multidirectional laminates. To characterize the impact of the variability of the properties of the base ply on the structural properties, uncertainties of the meso-scale (the ply) can be propagated to the macro-scale (the laminate and the structure). Besides the mechanical interest, such practice is interesting as it can to as certain extent (in)validate the assumptions on which the calibration process is based. For instance, to ensure the appliability of the method, though it is questionable, it has been hypothetized that the properties of all plies for a given laminate were identical. With uncertainty propagation, the impact of such an assumption on the variability of the material properties can be tested.
Failure criteria are one of the most important design criteria to ensure the safety of structures. Remember that with the standard propagation method (Monte-Carlo based methods [Metropolis and Ulam, 1949]), the quantities to be approximated require many samples to converge to their exact estimates, which demands to run the behavior law as many times as possible. To avoid a useless increase of the complexity of the propagation (requiring for instance to set up a surrogate model [Forrester et al., 2008]), it is necessary to strike a balance between the complexity of the propagation test-case (the more complex the more interesting) and the computational costs (the more complex the more costly). Determining the fracture stress of a perforated composite plate is one available test case that provides a compromise between these two requirements. Indeed, if we restrain ourselves to a linear elastic behavior, it is possible to use the Tan model [Tan, 2017], an analytical model which provides, for a given a load profile, the stress distribution over the hole plate. The failure stress can be then determined for example using the Hashin criterion [Hashin, 1980] combined with a point stress method. In the following, the different propagation steps are first described in Section 4.7.1. Then, the propagation results are analyzed for different laminates to investigate the impact of different assumptions and compare the impact of the calibrated distributions in Section 4.7.2.

4.7.1 Description of the propagation test-case

The objective of the uncertainty propagation is to estimate the empirical distribution of the failure stress of a perforated plate of composite material in tension. The propagation method is the Monte-Carlo method [Metropolis and Ulam, 1949]. The variables considered as random are the compliances $S_{11}^0, S_{22}, \nu_{12}$ and S_{66}. All other settings (geometrical features, resistances) are kept deterministic. To approximate this distribution, samples of failure stress should be generated, which implies to perform 3 different calculation steps. Remember that the calibrated distributions characterize the behavior of the ply from elementary laminates, that is to say UD 0°, UD 90° and ±45°. However, if they are suited for calibration because they emphasize specific aspects of the material behavior, their behaviors are too specific to be used within structures. In fact, more frequently, more complex laminates that mix different orientations are preferred, such as quasi-isotropic laminates which can mix $0°, 90°$ and $\pm 45°$ plies. Homogenization techniques should be employed to characterize the behavior of such laminates. The reference method to estimate the linear elastic properties of a laminate given a stacking sequence (the ordered list of the orientation and thickness of the plies in the laminate) and the elastic linear properties of the base ply (those characterized in this chapter) is the Classical Laminated Theory (CLT) from [Berthelot, 2012]. From the homogenized linear behavior it is possible to estimate the stress field over the perforated plate for a given applied load with the Tan model provided its geometric characteristics (width, length, radius of the hole). Finally, with the Hashin criterion and a Point Stress method, the failure stress can be determined. The different steps of the failure analysis, detailed further below, are summed up in Figure 4.31.

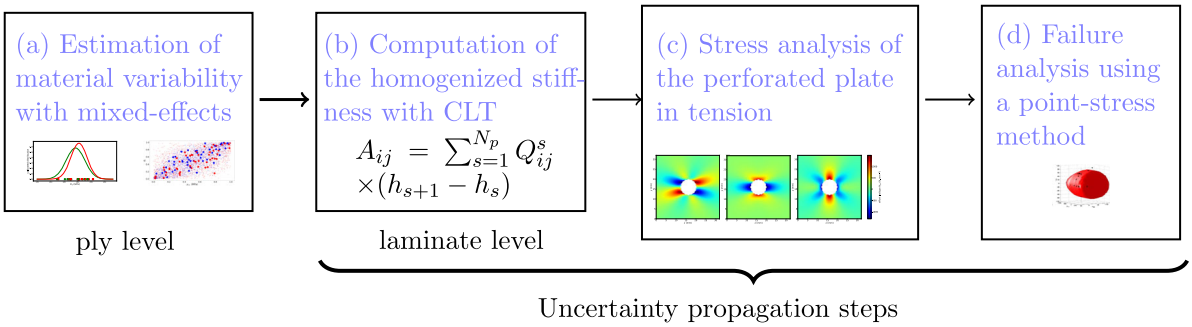


Figure 4.31: Main steps of the open hole plate in tension (OHT) failure analysis. After the calibration of joint distribution of the base ply properties (a), the elastic behavior of the laminate is computed with the CLT [Berthelot, 2012] (b). Thanks to the Tan model [Tan, 2017], the stress field for a perforated plate is approximated (c). Finally, a failure criterion [Hashin, 1980] is computed (d).

The Tan model is only valid if the behavior of the material is elastic linear. Yet, remember that in the calibration process, a more general nonlinear model has been considered. It is therefore necessary to extract from the joint distribution of $S_{11}^T, S_{11}^C, S_{11}^0$ and ν_{12} the joint distribution of S_{11}^0 and ν_{12} (the elastic linear properties). As only independent properties are considered, this is tantamount here to extract the marginals of S_{11}^0 (simplified into S_{11} in the following) and ν_{12}. Those of S_{22} and S_{66} calibrated separately are added to the marginals of S_{11} and ν_{12} and the joint distribution of the elastic properties simply reads as the product of all marginals. The compliance matrix $\underline{\underline{\mathbf{S}}}$ can be defined from these properties. However, this matrix is defined from S_{12} and not ν_{12}, that expresses $S_{12} := -\nu_{12} S_{11}$. In practice, Poisson's ratio ν_{12} and longitudinal stiffnesses S_{11} are generated from the calibrated distributions, before computing S_{12}. It should be noted that properties of this matrix are supposed to be all positive except S_{12} (which is in general negative). Thus, the calibrated Gaussian distributions that have infinite support are truncated so that no negative sample is generated. This is sufficient to ensure the fact that S_{12} remains negative. In addition, the compliance matrix $\underline{\underline{\mathbf{S}}}$ should be positive-definite [Berthelot, 2012]. For the sampling, the generated elastic properties that do not comply with this requirement are not

considered for further calculations. In total, this makes three distributions to be propagated, one for each strategy of the calibration ($\mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{C}_3$). It should be noticed that most of the parameters drawn from the calibrated distributions comply with these requirements, justifying *a posteriori* why these constraints were not added to the calibration process as it would have resulted into unnecessary complications.

4.7.1.1 Classical laminate Theory (CLT)

The composite materials used to build structures often appear as laminates [Berthelot, 2012], that is to say as a combination of base plies. They can be characterized by stacking sequences composed of a number of plies N_p, their associated orientations $\Upsilon_s \ \forall s \in [\![1, N_p]\!]$ and associated thickness $e_s \ \forall s \in [\![1, N_p]\!]$. The notions of orientation and thickness are illustrated in Figure 4.32.

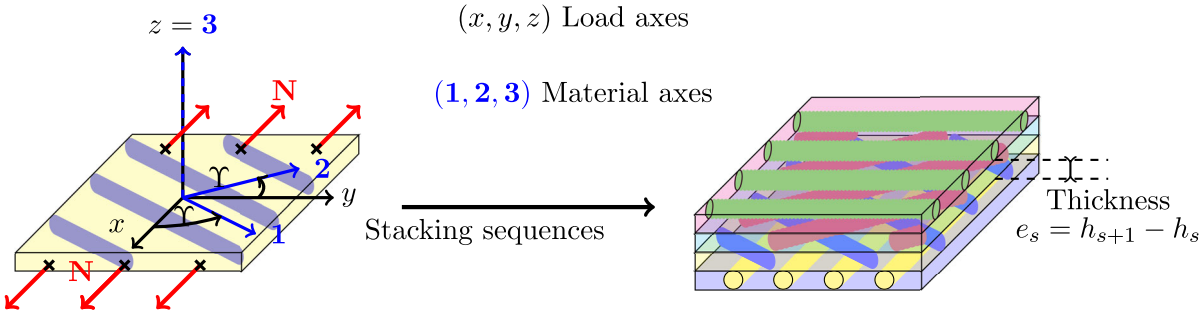


Figure 4.32: On the left, load and material axes alongside the base ply. On the right, $[0°, \Upsilon, -\Upsilon, 90°]$ laminate with the thickness illustration.

Figure 4.32 presents the geometrical features of the base ply, in particular the notion of orientation. The vector of the generalized load $\mathbf{N} = (N_x, N_y, N_z)$ is defined with respect to the global frame (x, y, z). The orientation of the plies Υ_s refers to the angle between the fiber and the $x-$direction. A laminate $[0°, \Upsilon, -\Upsilon, 90°]$ is illustrated in Figure 4.32. The thickness of the s^{th} ply e_s can be also defined as the difference of the height (along the z-axis) of the upper face of the ply h_{s+1} and the height of its lower face h_s.

The CLT is an homogenization method that allows to compute the macroscopic behavior of a laminated composite. It relies on the Kirchoff-Love assumptions (see [Berthelot, 2012] for more details). The constitutive equation of the laminate relates the generalized in-plane load N and bending load M to the mid-plane strain ε^0 and the curvature κ:

$$\begin{pmatrix} N \\ M \end{pmatrix} = \begin{pmatrix} A & B \\ B & D \end{pmatrix} \begin{pmatrix} \varepsilon^0 \\ \kappa \end{pmatrix} \ . \tag{4.28}$$

A, B and D express as weighted sums combining the elastic and geometric characteristics of the plies (orientations $[\Upsilon_s]_{s\in[\![1,N_p]\!]}$ and heights $[h_s]_{s\in[\![1,N_p]\!]}$). Only membrane load will be considered in the following, so the D (bending) matrix is not considered further. For a laminate with N_p plies, the A and B matrices are expressed as:

$$A_{ij} = \sum_{s=1}^{N_p} Q_{ij}^s (h_{s+1} - h_s) \ , \ B_{ij} = \frac{1}{2} \sum_{s=1}^{N_p} Q_{ij}^s (h_{s+1}^2 - h_s^2) \ , \tag{4.29}$$

with Q_{ij}^s coefficient (i, j) of the reduced stiffness matrix (the inverse of the compliance matrix $\underline{\underline{\mathbf{S}}}$ in plane stress) of the s^{th} ply in its global frame. Note that the B matrix is zero for symmetric lay-ups, which will not be necessary the case here.

Aside from the elementary laminates (0°,90°,±45°), two stacking sequences corresponding to quasi-isotropic laminates, widely used in industry, will be processed: $[0°, 90°, -45°, 45°]_s$ labeled QI_1 and $[0°, 60°, -60°]_s$ labeled QI_2 (index s standing for symmetric). Each type of laminate will be considered of unitary thickness, so the thickness of each ply corresponds to $\frac{1}{N_p}$.
Remember that in the calibration process, all plies were described by the same realization of the elastic constants. If this choice was necessary because of applicability matters, those one do not stand anymore here. Moreover, assigning to each ply different sets of elastic parameters seems more realistic from a physical point of view. Indeed, in the manufacturing process, base plies are cut independently from the others within the same sheet of composite material in which material variability expresses. To study this particular point, for 0° and 90° laminates, the propagation is carried out considering first the same elastic properties for all plies (similar to consider only one ply), and then laminates with 4 and 8 plies whose elastic properties are different from one another. For the other laminates, all plies are assigned different values for the elastic properties. In the cases where the plies are assigned different properties, the samples characterizing the material properties of each ply are independent identically distributed (i.i.d.). This gives us 9 laminates for which uncertainty propagation is achieved. These cases are reported in Table 4.22.

Table 4.22: Tested laminates.

Scenario	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_4$	$\mathcal{S}_5$	$\mathcal{S}_6$	$\mathcal{S}_7$	$\mathcal{S}_8$	$\mathcal{S}_9$
Orientations	0°	0°	0°	90°	90°	90°	±45°	$[0°, 90°, \pm 45°]_s$	$[0°, \pm 60°]_s$
Number of plies	1	4	8	1	4	8	4	8	6

For the scenarios where the properties of the plies are different from one another, the final laminate is no longer symmetric, and therefore the B matrix is non-zero, making calculations harder to perform. Yet, its contribution may be negligible with respect to the one of the A matrix. To evaluate this point, for 10,000 realizations of the elastic constants, it is possible to estimate the ratio γ defined as the maximum absolute value of the B matrix to its counterpart of the A matrix. Still, as they are not of the same dimension (the A matrix is expressed in MPa.m and the B is expressed in MPa.m^2 with m referring to meters), the ratio to study is expressed as $\gamma = \frac{2B_{\max}}{eA_{\max}}$ [Julien, 2010]. $A_{\max}$ and $B_{\max}$ respectively refer to $\max_{(i,j)\in[\![1,3]\!]^2} |A_{ij}|$ and $\max_{(i,j)\in[\![1,3]\!]^2} |B_{ij}|$, with A_{ij} and B_{ij} defined in Eq. (4.29). The maximum is taken with respect to all components of the matrices because the presence of the hole within the plate creates shear stresses.

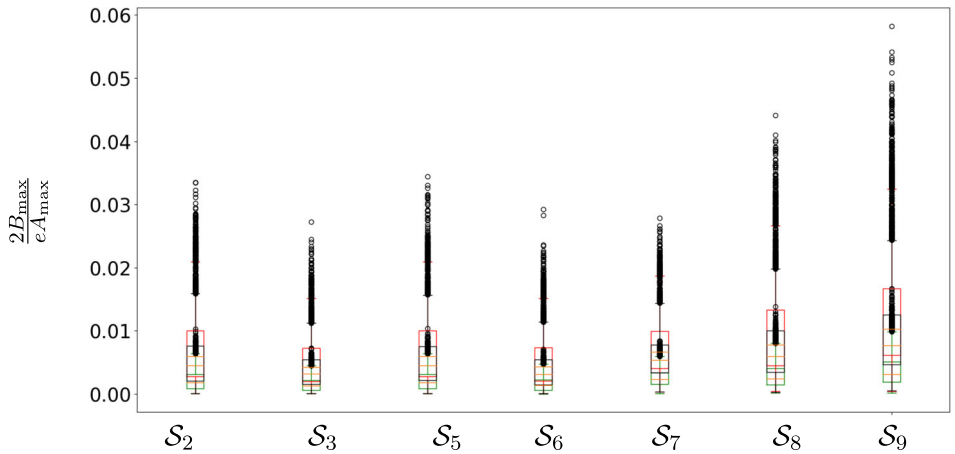


Figure 4.33: Gamma ratio for several scenarios for 10,000 realizations

In Figure 4.33, for all scenarios, the contribution of the B matrix to overall behavior seems negligible with respect to the contribution of the A matrix. Thus, even if the laminates are not

perfectly symmetric, neglecting the B matrix appears as reasonable approximation and will be the case in the following. Note also that, though this indicator is simple to use, it has a major drawback. Indeed, it assumes that the magnitude of the deformation of the mid-plane and the curvature are comparable, which is difficult to interpret from a physical point of view as those quantities represent different physical phenomena. Therefore, even if in this work no further analyses will be made because we do not aim to size structures, in practical cases, additional testings should be conducted with more advanced tools.

4.7.1.2 Stress analysis with the Tan Model

The Tan Model is an analytical model which enables to estimate the macroscopic stress field of a perforated plate of composite material. Before proceeding further into the analysis of the Tan model, let's present the perforated plate in tension along with the relevant quantities.

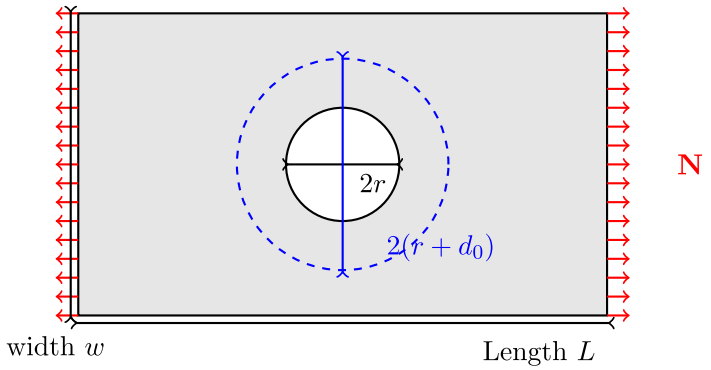


Figure 4.34: Perforated plate in tension under load **N**, r stands for the radius of the hole and d_0 stand for the point-stress distance discussed next section.

In Figure 4.34, the key features of the perforated plate in tension are illustrated. A macroscopic longitudinal load F is applied to both ends of the plate of width w and length L. The hole with radius r is located in the center of the plate.
The Tan model is an analytical model which allows to avoid the use of Finite Element Analysis that is computationally expensive. It provides the stress field over the plate excluded from the circle determined by the radius and the point-stress distance. It relies on two main assumptions:

- The plate is infinite;
- The behavior of the material is orthotropic elastic linear.

Of course, in practice, the first assumption is not realistic, requiring to add a correction factor to the approximated stresses by the Tan model. It is equal to

$$C_{wd} = \frac{2 + (1 - r_{wd})^3}{3(1 - r_{wd})} \ , \tag{4.30}$$

with $r_{wd} = \frac{2r}{w}$ where r the radius of the hole and w the width of the plate. This expression is specific to quasi-isotropic lay-ups and it is supposed to be valid for all other lay-ups. The second hypothesis explains why we focus in this part on the linear elastic behavior by extracting marginals of the relevant parameters (that is to say $S_{11}, \nu_{12}, S_{22}, S_{66}$). An analysis similar to the one conducted in Figure 4.33 shows that both A_{16} and A_{26} are negligible with respect to the other terms of the A matrix in all scenarios, and thus, the behavior of all laminates is orthotropic elastic linear. The geometric characteristics employed here are reported in Table 4.23 and refer to usual values. Note that the length L is not involved in the calculation of the Tan model.

Table 4.23: Main characteristics of the perforated plate. All quantities are deterministic.

Width w [mm]	Point-stress distance d_0 [mm]	Radius of the hole r [mm]	Load N_{xx} [N.mm^{-1}]
32.0	0.710	3.18	500

For quasi-isotropic laminate QI_1, the execution of the Tan model gives the following estimation of the macroscopic stress fields for the longitudinal, transverse and shear component in Figure 4.35.

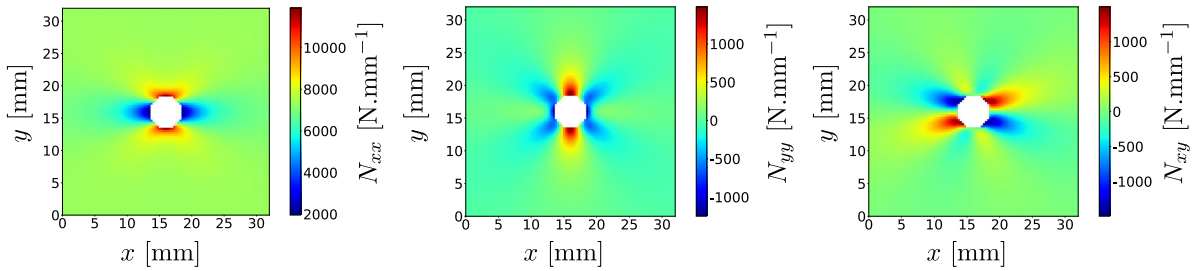


Figure 4.35: For a quasi-isotropic laminate QI_1, estimation of the macroscopic stress field with the Tan model. On the left, the longitudinal component; in the middle, the transverse component; on the right, the shear component.

Once the macroscopic stresses are estimated, the objective is to estimate the failure stress of the laminate, which refers to the lowest of the failure stresses of each ply. It should therefore be estimated for each ply, demanding to localize the macroscopic stress in each ply. In the CLT framework, the strain is assumed to be the same accross the plies, equal to ε^0 that can be computed *via* Eq. (4.28). Then, in the appropriate frame, the local stress is obtained from Eq. (4.14).

4.7.1.3 Failure analysis : Point-Stress and Hashin criterion

Around the hole boundary, an elastic behavior leads to infinite stresses on the hole boundary, which is not physically relevant. A more realistic modeling should include other phenomena such as damage which would avoid such situations. If we want to keep using with the elastic model, one possibility can be to increase the distance between the center of the hole and the point at which the failure criterion is applied. This method is called "point-stress" method [Nuismer and Whitney, 1975]. It consists in neglecting the phenomena that occur around the hole. The size of this region is characterized by the point-stress distance d_0. The determination of d_0 is semi-empirical as it relies both on experiments and Finite Element Analysis. In theory, it depends of the considered laminate but for the sake of simplicity, it will be the same for all laminates. In addition, as the highest stresses will be found (and thus where failure will occur) on the edges of the hole of radius $r + d_0$, during propagation, the stress is computed only on the edge of the hole of radius $r + d_0$ to reduce computational costs.
Failure itself will be studied with the Hashin criterion [Hashin, 1980], a widely used failure criterion that helps to predict the failure of a composite ply. From the local stresses, the idea is to compare the quantities below to one. If it is greater or equal to 1, failure occurs. Two failure mode corresponding to the constituents (fiber and matrix) are considered. The quantities $\mathcal{Q}(\sigma_{11}, \sigma_{22}, \sigma_{12}, \mathcal{Y})$ to be compared to the threshold stand for:

$$\begin{cases} \left(\dfrac{\sigma_{11}}{X_t}\right)^2 \text{ in tension, } \left(\dfrac{\sigma_{11}}{X_c}\right)^2 + \left(\dfrac{\sigma_{12}}{S_c}\right)^2 \text{ in compression,} & \text{(4.31a)} \\ \left(\dfrac{\sigma_{22}}{Y_t}\right)^2 + \left(\dfrac{\sigma_{12}}{S_c}\right)^2 \text{ in tension, } \left(\dfrac{\sigma_{22}}{Y_c}\right)^2 + \left(\dfrac{\sigma_{12}}{S_c}\right)^2 \text{ in compression,} & \text{(4.31b)} \end{cases}$$

where $\mathcal{Y} := [X_t, X_c, Y_t, Y_c, S_c]$ with X_t the fiber failure stress in tension, X_c the fiber failure stress in compression, Y_t the matrix failure stress in tension, Y_c the matrix failure stress in compression

and S_c the shear failure stress (all in MPa). The criteria of the first line refer to the failure of the fiber and those of the second line to the failure of the matrix.
The Hashin criterion does not provide the value of the failure stress, it just predicts the occurrence of failure for given load value . To estimate the failure stress for any configuration of the material properties, it is possible to compute the reserve factor that can be defined as the coefficient by which the macroscopic load should be multiplied to reach failure. This comes from the fact that the relation between the macroscopic load and the local stresses in each ply (see Eq. (4.28)) is linear. The reserve factor is finally expressed as $\frac{1}{\sqrt{\mathcal{Q}(\sigma_{11},\sigma_{22},\sigma_{12},\mathcal{Y})}}$. In scenarios for which the laminates are composed of more than one ply, the reserve factor is determined for all plies and the reserve factor of the laminate corresponds to the lowest value. The failure properties refer to those of T700GC/M21 material detailed in [Laurin, 2005] and are reported in Table 4.24.

Table 4.24: Failure properties from T700GC/M21 material.

X_t [MPa]	X_c [MPa]	Y_t [MPa]	Y_c [MPa]	S_c [MPa]
1.90×10^3	-1.00×10^3	84.0	-250	75.0

The failures of open hole plates of 0° and 90° laminates in tension are mainly driven by the matrix. To understand this point, it is necessary to remember that for polymer matrix composites, the resistances associated to the matrix (Y_t, Y_c and S_c) are much lower than the resistances associated to the fiber (X_t and X_c, see Table 4.24) and thus, the cracks which may lead to failure will appear first within the matrix. Contrary to multi-orientation laminates, the cracks that appear in one ply cannot be stopped by the fibers of the upper/lower plies, leading to the final failure. In fact, for such stacking sequences, the relevant failure is the matrix failure criterion (in scenarios $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4, \mathcal{S}_5$ and $\mathcal{S}_6$). This same reasoning leads us to consider for multi-orientation laminates (scenarios $\mathcal{S}_7, \mathcal{S}_8$ and $\mathcal{S}_9$) the fiber failure criterion.

4.7.2 Propagation results

The propagation results are now presented. We start with the analysis of the distribution of the failure stress for the elementary laminates (0°, 90° and ±45°) to investigate the key points before moving onto the quasi-isotropic laminates that are of practical interest in design problems. For each scenario, 10,000 samples of elastic properties have been drawn from each calibrated distribution. Next, for each of these elastic properties, the corresponding failure stress is estimated after completion of the stages described in Figure 4.31. For each laminate, the empirical distributions of the samples of failure stress are estimated with Kernel Density Estimation [Wand and Jones, 1994] implemented in OpenTURNS [Baudin et al., 2016]. The red, green and black colors respectively refer the calibration scenarios ($\mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{C}_3$) defined in the first paragraph of Section 4.5.5. The distributions characterizing longitudinal elasticity are combined to those of two other statistically independent variables, S_{22} and S_{66}.

4.7.2.1 Examination with elementary laminates

To understand the propagation behavior, we first focus on the elementary laminates that were processed to calibrate the elastic model. Those one refer to 0°, 90° and ±45° laminates. To start the analysis, we concentrate on the unidirectional laminates (0° and 90° laminates). The first point that ought to be checked is the impact of the input distribution of the material properties on the distributions of the failure stress. Indeed, as it can be noticed in Figure 4.30, the standard-deviation of S_{11} and of S_{12} (through ν_{12}) can have very different values depending on the calibration strategy. For scenarios $\mathcal{S}_1$ and $\mathcal{S}_3$, the empirical distributions of the (matrix) failure stress are depicted in Figure 4.36.

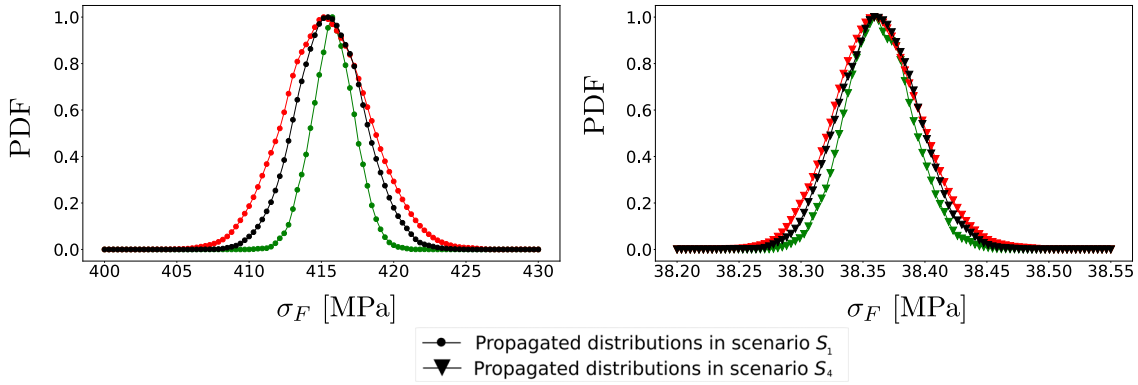


Figure 4.36: PDF of the empirical distributions estimated by KDE from 10,000 samples of failure criterion σ_F for scenario $\mathcal{S}_1$ (on the left) and $\mathcal{S}_4$ (on the right). σ_F refers to the failure stress. The red, green, and black colors respectively refer the calibration scenarios ($\mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{C}_3$) defined in the first paragraph of Section 4.5.5.

A first feature that can be noticed in Figure 4.36 is that the size of the support of the three distributions of the failure stress seem to follow the same trend as the standard-deviations of the longitudinal linear compliance S_{11}^0, in particular for the 0° laminates (left of Figure 4.36). Indeed, as it can be spotted in Figure 4.30 and reported in Tables 4.18, 4.17 and 4.20, the standard-deviation of the longitudinal linear stiffness S_{11} calibrated in case $\mathcal{C}_1$ (depicted in red Figure 4.36) is larger than the standard-deviation of S_{11} estimated in case $\mathcal{C}_3$ (depicted in green in Figure 4.36), itself larger than the standard-deviation of S_{11} in case $\mathcal{C}_2$. This behavior can be explained by matters of sensitivity: the response of 0° laminates loaded in the fiber axis corresponds approximately to the fiber behavior characterized by S_{11}. In other words, for such laminates, the main contribution to the elastic output originates from this parameter, explaining the high sensitivity to the variability of this parameter of the failure stress. On the contrary, as the effect of S_{12} which accounts for the impact of the Poisson's ratio is negligible on the model output compared to S_{11}, the impact of the variation the standard-deviation of ν_{12} which follows a different pattern compared to the one of S_{11}^0 is almost impossible to notice on the PDF curves. A comparable pattern seems to arise for UD 90° laminates. However, it is harder to draw clear conclusions because the range of the support of the failure stress distribution is much smaller for 90° laminates than for 0° laminates. Therefore, the differences could be more likely the consequences of the limited number of samples and of numerical errors within the KDE phase. These limited changes between the distributions can be again explained by sensitivity issues as the main contribution to the behavior of such a laminate comes from the matrix response characterized by S_{22} and to a smaller extent S_{12}. This may partly explain the small changes between the distributions, as the distribution of the transverse compliance remains always the same. It can be also observed that the failure stress for the 90° laminates is much smaller than for 0° laminates. This is consistent with the fact that the response of 90° laminates loaded in the longitudinal axis is monitored by the matrix whose resistance is smaller compared to that of the fiber, as reported in Table 4.24.

Another aspect that can be investigated is the impact of the number of plies on the variability of the failure stress. Remember that, in the calibration process, it has been chosen to consider that all plies within the elementary laminates used for the identification were characterized by the same elastic properties. This assumption can be challenged. Indeed, in the manufacturing process, the plies used to build the final laminate can be taken from different composite sheets that can be considered as different realizations of the material properties. To examine this particular point, it is possible to take a look to the empirical distributions of the failure criterion in scenarios $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ for UD 90° laminate illustrated in Figure 4.37.

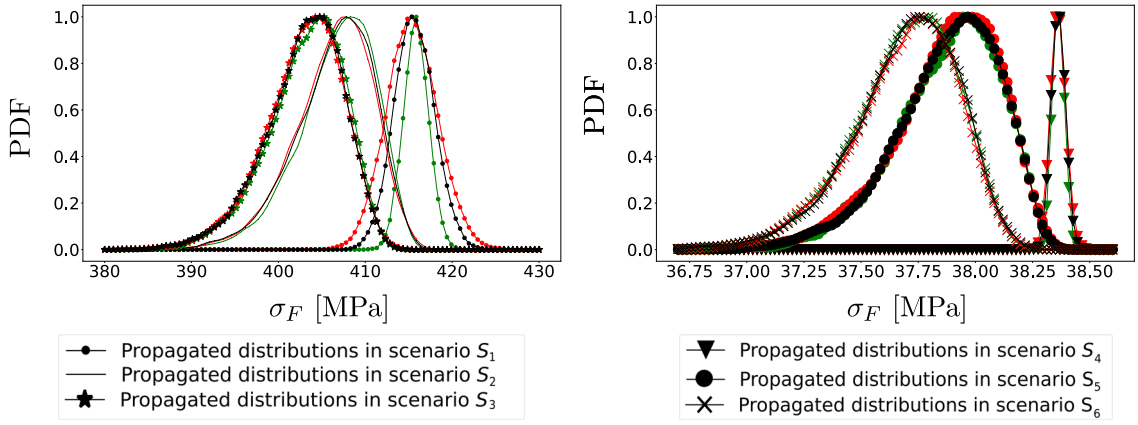


Figure 4.37: PDF of the empirical distributions estimated by KDE from 10,000 samples of failure criterion σ_F for scenarios $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ (on the left) and $\mathcal{S}_4$, $\mathcal{S}_5$ and $\mathcal{S}_6$ (on the right). σ_F refers to the failure stress. The red, green, and black colors respectively refer the calibration scenarios ($\mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{C}_3$) defined in the first paragraph of Section 4.5.5.

In Figure 4.37, it can be noticed that the number of plies has an impact on the dispersion of the failure stress. In particular, it shows that in average, the failure stress decreases as the the number of ply increases which can be explained numerically. Indeed, as the number of plies within a laminate increases, each ply can lead to failure. As there is a higher probability for one of the plies to be poorly resistant, the failure of the laminate is hastened. On the contrary, with only one ply, the laminate only reaches failure when the unique ply breaks, which limits the risk of failure and increase the value of the failure stress. However, this analysis should be considered cautiously from a physical point of view as several simplifications have been made in the modeling of the material behavior and more complex analyses should be performed to provide more reliable conclusions. For instance, the resistances have been taken as deterministic while these properties are known to be subjected to a significant level of variability. Taking such aspect into account could stand as future works on this topic. At least, it highlights the importance to test the impact of assigning different properties to the plies of a laminate. In addition, it increases the variability of the failure criterion.

This point deserves further inquiries as this result is not obvious. On the one hand, for a laminate with N_p plies, N_p sets of elastic properties are drawn instead of only one for a single ply laminate, enlarging the possible configurations for the laminates and thus the possible values for the failure stress. On the other hand, the propagation includes an homogenization step (the CLT) that averages the variability of the mechanical properties across the laminate. Therefore, we aim to study the impact of the variability of the mechanical properties in the two steps where they can influence the propagation results: the calculation of the membrane matrix (A matrix) and the estimation of the failure stress which requires to estimate the local stresses in each ply. Note that this point is of major concern because it could invalidate to certain extent the assumptions on which the calibration method is based. To investigate the impact of the variability of the mechanical properties on the membrane matrix, let's take the example of the longitudinal stiffness A_{xx}. To illustrate this impact, for all possible observation orientations within $[0°, 360°]$, the A matrix for the laminates of scenarios $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4, \mathcal{S}_5$ and $\mathcal{S}_6$ is evaluated. This process is repeated 500 times for each laminates and provides the polar diagrams proposed in Figure 4.38.

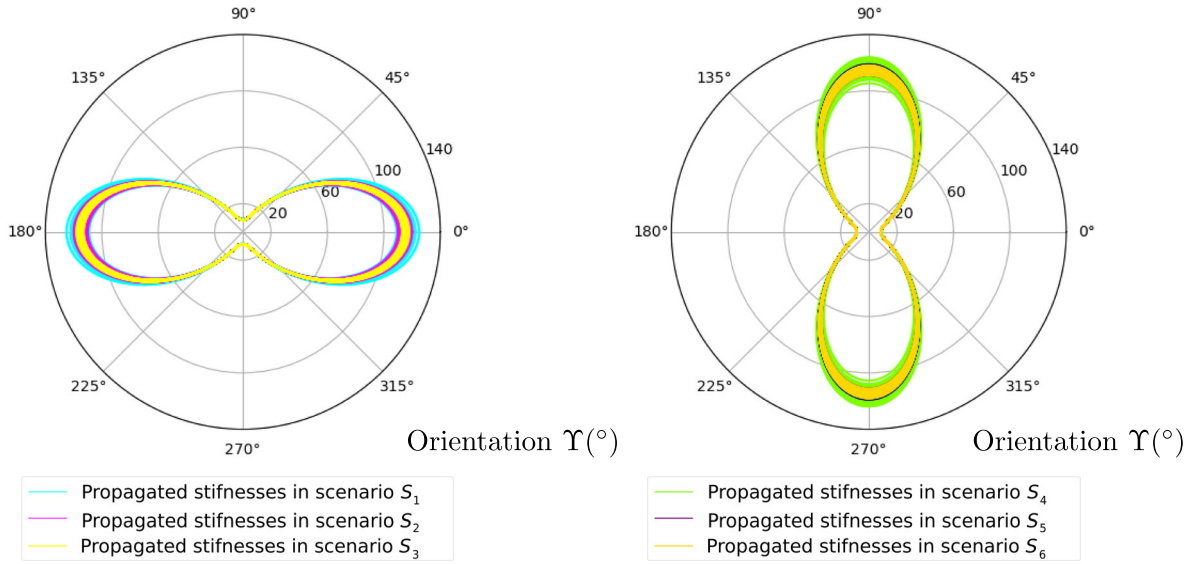


Figure 4.38: On the left, polar plot of A_{xx} (in GPa) for laminates of scenarios $\mathcal{S}_1, \mathcal{S}_2$ and $\mathcal{S}_3$ for 500 realizations of the elastic constants drawn from the distribution calibrated in case $\mathcal{C}_3$. On the right, polar plot of A_{xx} (in GPa) for laminates of scenarios $\mathcal{S}_4, \mathcal{S}_5$ and $\mathcal{S}_6$ for 500 realizations of the elastic constants drawn from the distribution calibrated in case $\mathcal{C}_3$. The radial axis refers to the value of A_{xx} in GPa.

Figure 4.38 clearly suggests that the computation of the A matrix is not the step that increases the variability of the failure stress. Indeed, the laminates for which the stiffness is the most variable are the ones with only one ply. On the contrary, with several plies, the variability of A_{11} decreases, as if the contribution of the different plies were averaged. In fact, if we want the variability of A_{11} in scenarios $\mathcal{S}_2$ and $\mathcal{S}_3$ to reach the same magnitude as in scenario $\mathcal{S}_1$, it is necessary to consider wider dispersions for the elastic constants. This could suggest that the standard-deviations calibrated under the hypothesis of a single set of elastic constants for the whole laminate are underestimated, given that behind the calibration procedure of the distribution of the elastic constant with the elementary laminates discussed in Section 4.2.1, there is an implicit implementation of the CLT. This remark is of prime importance with respect to our concern about the assumption of considering the same set of elastic properties for the all plies in the calibration phase and would deserve to be investigated further in a dedicated study.

The other step in which the variability of the elastic properties can impact the propagation results is the estimation of the local stresses in each ply. From these local stresses the failure stress is computed for each ply and the failure stress of the laminate is in fact the lowest of all criteria. To examine this point, it can be interesting to take a look to the distribution of the index of the broken ply. The histograms of this (discrete) distribution are depicted in Figure 4.39 for UD 0° in scenarios $\mathcal{S}_2$ and $\mathcal{S}_3$.

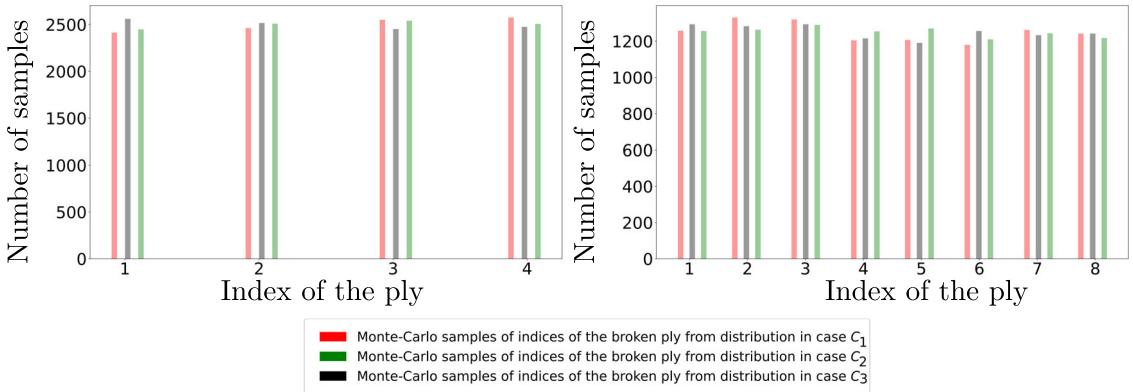


Figure 4.39: For scenarios $\mathcal{S}_2$ and $\mathcal{S}_3$, histogram of the discrete distribution of the index of the broken ply for 10,000 laminates.

Figure 4.39 shows that all ply break with equal probability, irrespective of the number of plies in the laminate, whatever distribution is considered. In fact, given that all plies are assigned specific elastic properties, as the failure criterion are estimated independently on the different plies, the effective dimension of the underlying distribution of the elastic properties is not only 4 (the number of elastic properties) but rather $4 \times N_p$. This is the decisive factor that increases the variability of the failure criterion and underlines the interest of such examinations.
To terminate this study, let's take a look to the empirical distributions of the failure criterion with the $\pm 45°$ laminate (scenario $\mathcal{S}_7$) depicted in Figure 4.40.

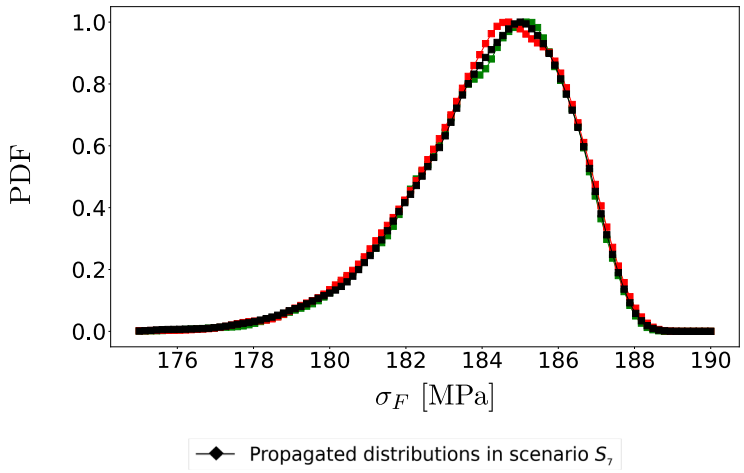


Figure 4.40: PDF of the empirical distributions estimated by KDE from 10,000 samples of failure stress σ_F for scenarios $\mathcal{S}_7$ (on the right). σ_F refers to the failure stress. The red, green, and black colors respectively refer the calibration scenarios ($\mathcal{C}_1$,$\mathcal{C}_2$ and $\mathcal{C}_3$) defined in the first paragraph of Section 4.5.5.

To start the analysis of Figure 4.40, it can be noticed that the empirical distributions of the failure stress for $\pm 45°$ are almost unchanged for the different calibrated distributions. This can be mainly explained by the negligible contribution of both S_{11} and S_{12} to the elastic behavior of $\pm 45°$. It can also be noticed that the predicted failure stresses for the $\pm 45°$ laminates are higher than those for the 90°. This is consistent from a physical point of view as the behavior of 90° laminates is monitored almost totally by the matrix when the behavior of $\pm 45°$ combines together that of the fiber and matrix. It is harder to make comments on the value of the failure stress for the 0° because the simplified modeling retained here does not take into account complex phenomena such as visco-elasticity,splitting [Jawaid et al., 2018], early failure of fibers which lower the resistance of the laminate.

4.7.2.2 Examination with quasi-isotropic laminates

This section is dedicated to the analysis of the propagation results for the quasi-isotropic laminates QI_1 and QI_2 laminates, corresponding to scenarios $\mathcal{S}_8$ and $\mathcal{S}_9$. The empirical distributions of the failure criterion are reported in Figure 4.41.

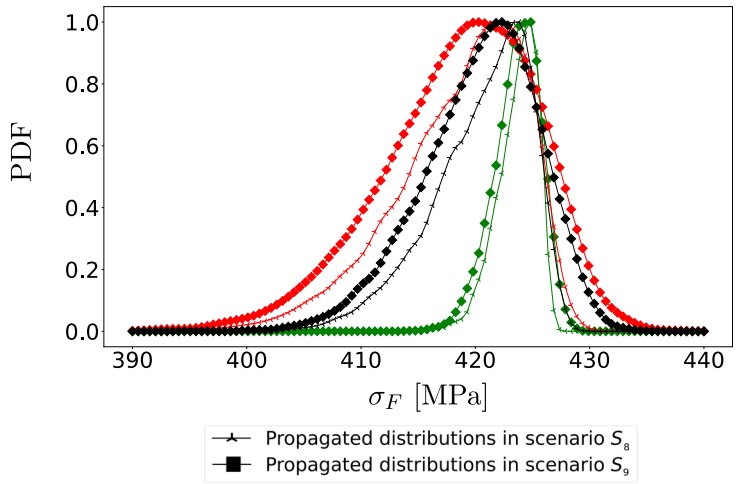


Figure 4.41: PDF of the empirical distributions estimated by KDE from 10,000 samples of failure stress σ_F for scenarios $\mathcal{S}_8$ and $\mathcal{S}_9$. σ_F refers to the failure stress. The red, green, and black colors respectively refer the calibration scenarios ($\mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{C}_3$) defined in the first paragraph of Section 4.5.5.

In Figure 4.41, it appears that the behavior of QI_1 and QI_2 laminates are significantly impacted by the change of distributions, which is expected because both of these laminates include 0° base plies in their stacking sequence (see 4.22). This also illustrates the interest to include within the laminate 0° and to a lesser extent 90° plies to strengthen it. Indeed, the average failure criterion for quasi-isotropic laminates is almost twice the average failure criterion for ±45° laminates (see Figure 4.40). It also shows that both laminates offer similar resistances. Furthermore, it can be noticed that the failure stresses are similar to those of the UD 0° laminates (see Figure 4.37). It is difficult to compare the values of the failure stress because physical phenomena are omitted for both types of laminates (splitting for the unidirectional laminates and viscoelasticity/damage for the quasi-isotropic ones). However, splitting can be thought as more severe as it separates the fibers from one another, which causes almost immediate failure. Furthermore, the assumption on the total thickness of the laminate supposed equal deterministic equal to one is strong. Indeed, in practice, at the beginning of the manufacturing process, only the thickness of the plies are known, while the thickness of the laminate can change from one batch to the other. Consequently, the values of the failure stress should be much lowered to get reliable predictions, but this is out of the scope of this work. To better discriminate against the two laminates, it would be necessary to include more complex phenomena that are not considered here.

Remember that the laminates corresponding to scenarios $\mathcal{S}_8$ and $\mathcal{S}_9$ are composed of plies of different orientations, and thus do not offer the same resistance to the macroscale load. Consequently, it can be interesting to examine for both laminates which ply breaks the first and provoke the final failure of the laminate as illustrated in Figure 4.42

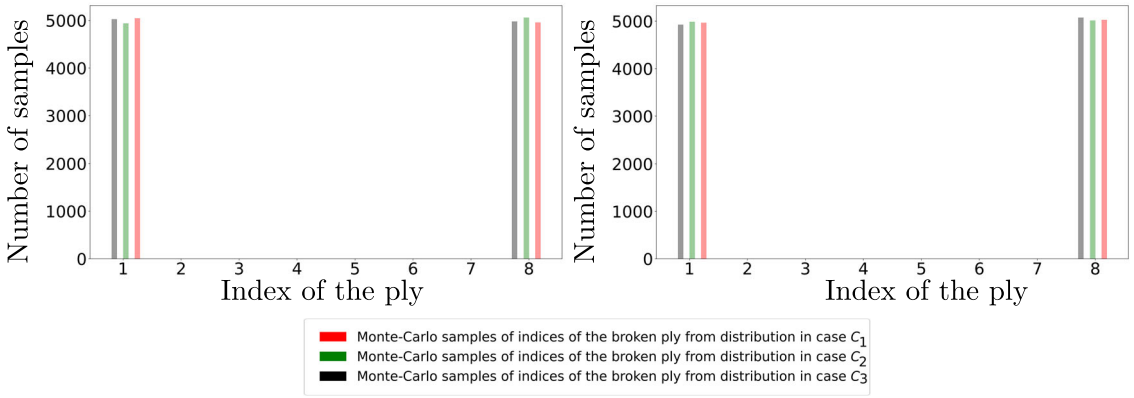


Figure 4.42: For scenarios $\mathcal{S}_7$ and $\mathcal{S}_8$, histogram of the discrete distribution of the index of the broken ply for 10,000 laminates.

Figure 4.42 shows that for all kinds of distribution and stacking sequences of quasi-isotropic laminates, it is always the same kind of ply that breaks within the laminates: the outer plies which refer to the 0° plies. Considering the simplistic behavior law, this is not surprising. Indeed, the failure criterion studied here is the fiber criterion. Consequently, the most loaded ply is the one that is aligned with the fiber direction, that is to say here the 0° plies.

This short study has allowed to illustrate a complete propagation process, from the calibration of the material variability. This process could be carried out again with more complex and representative models.

4.8 Conclusion

This chapter presents a sequential calibration procedure that aims to calibrate complex material behavior laws compliant with the mixed-effects framework using multivariate data. The method consists in defining a sequence of calibration subproblems, each one corresponding to a subset of the available experimental data. The subdivision of the complete problem into subproblems can be based on the choice of the expert or on sensitivity analyses. Indeed, well-chosen tests can activate different parts of the model, *i.e.*, different subsets of the model parameters. Thus, each type of test makes it possible to define a subproblem whose objective is to identify the joint distribution of a subset of the model parameters. A parameter can be involved in several subproblems. In this case, it is identified a first time and then, the search space of the corresponding distribution parameters is reduced to a trust region in the subsequent calibrations, to ensure consistency between the different subproblems.

The method is applied to a virtual test case to calibrate an orthotropic elastic model in plane stress with nonlinear longitudinal behavior. This test-case is interesting as it allows to compare the sequential strategy to its direct counterpart using design matrices, what would not necessarily be possible in other experiments. A first implementation enables to split the identification of the joint distribution of the compliances with the usual elementary UD 0°, UD 90° and ±45° laminates. With the second, it was possible to divide the calibration of the joint distribution of the longitudinal and transverse nonlinear elasticity of UD 0° laminates between the tension and compression experiments. The consistency of the results throughout the calibration process is ensured by bounding the parameters estimated twice. The results of the methodology applied to synthetic data generated with the exact model demonstrate the ability of the proposed procedure to estimate properly the model parameters distribution in the presence of significant material variability. The results obtained considering the different orders of calibration are investigated and compared to the straightforward calibration of the complete joint distribution using design matrices. In particular, it shows that calibration with design matrices is not necessarily the best in our case, while it processes the highest amount of data. This is because of the trade-offs made between the different types of experiments. The application to a material with carbon and epoxy resin confirm this point and shows in particular that with a sequential strategy, the full estimation of the joint distribution takes, in our application (*i.e.*, model and data) about half the time of the calibration of the joint distribution. This is a key feature, and combined to the preservation of the calibrated distributions throughout the calibration process, it shows that sequential calibration is not only compatible with mixed-effects models but scales better to more complex model with a high number of parameters to be calibrated. Finally, we have propagated the calibrated distributions to estimate the failure stress of a perforated plate for different stacking sequences. This application has provided us the opportunity to test the impact of the different assumptions that have been made to achieve the calibration of the model parameters distribution, in particular concerning the impact of the variability on the plies properties. The propagation results have been confronted to physical observations in order to

check the consistency of the results, even if the chosen modeling here is too limited to draw clear conclusions on the different phenomena that occur within the laminates. This chapter has allowed us to explore all steps of uncertainty quantification from the estimation of material variability to its propagation in a structural test-case.
Still, from these studies several limitations can be pointed out. For instance, as in the previous chapter, model inadequacy (labeled $\boldsymbol{\zeta}$ in Chapter 1) has been shown to be significant with respect to the impact of measurements errors and its modeling in future inquiries may help to improve the calibration process. Next step could consist in modeling bias to better represent physical phenomena and comply with hypotheses of the statistical modeling. Furthermore, this study has been carried out on a simple material model whose execution is fast and it is possible to carry out many model evaluations. Even if it is beyond the scope of this work, it could be interesting to apply the sequential strategy to models that present stronger nonlinearities such as viscosity or damage. Given that such models are more computationally intensive than the nonlinear elasticity, it would allow to test the applicability of the method to complex behavior laws. Another limitation of this investigation is that with the tested model, standard mixed-effects can be applied. However, in practice, this may not be always the case. A preliminary study on this subject has been performed in [Laboulfie et al., 2022] but it deserves to be continued further to draw solid conclusions. Finally, it could be interesting to design methods to choose both β and α from statistics rather than arbitrarily. For instance, they could be chosen using Bayesian inference.

GENERAL CONCLUSION

Contributions

The contributions of this thesis have been the following:

- Talk: "Laboulfie, C., Balesdent, M., Brevault, L., Da Veiga, S., Irisarri, F.X., Le Riche, R. and Maire, J.F., 2021, June. Calibration of material model using mixed-effects models. In the 4th International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2021) [Laboulfie et al., 2021]

- Talk: and conference article: "Laboulfie, C., Balesdent, M., Brevault, L., Da Veiga, S., Irisarri, F.X., Le Riche, R. and Maire, J.F., 2022, August. Sequential calibration of material model using mixed-effects. In 25ème Congrès Français de la mécanique (CFM 2022) [Laboulfie et al., 2022]." Developed in a journal article "Laboulfie C., Balesdent M. , Brevault L., Maire J.-F., Da Veiga S., Le Riche R., Sequential calibration of material constitutive model using mixed-effects calibration, Mechanics & Industry **24**, 32 (2023) (2023) [Laboulfie et al., 2023] ".

- Talk: "Laboulfie, C., Balesdent, M., Brévault, L., Da Veiga, S., Irisarri, F.X., Le Riche, R. and Maire, J.F., 2022, June. Mixed-effects methods for material model calibration: influence of assumptions on parameter distributions. In the 8th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2022), Oslo, Norway"

Work Summary

The main objective of this thesis was to design a calibration method that is able to quantify the intrinsic variability of the material properties. For this purpose, the framework of the population approaches commonly used in pharmacokinetics has been introduced, in particular the mixed-effects models. An extension of the mixed-effects methodology, based on a sequential decomposition of the algorithm, has been proposed to handle the calibration of complex material behavior law. Finally, the uncertainty quantification process has been illustrated, starting with the characterization of the variability of the material properties and propagating them through a simulation model to estimate the failure stress of a perforated plate in tension. Our work shows the interest of using such mixed-effects approaches in material science. The different steps of the thesis are summarized in Figure 4.43.
The first step of the work, focused on model calibration, consists in exposing the objectives and the basics of this research field. A thorough literature review of the standard calibration methods, especially in the context of material behavior laws, has therefore opened this manuscript.

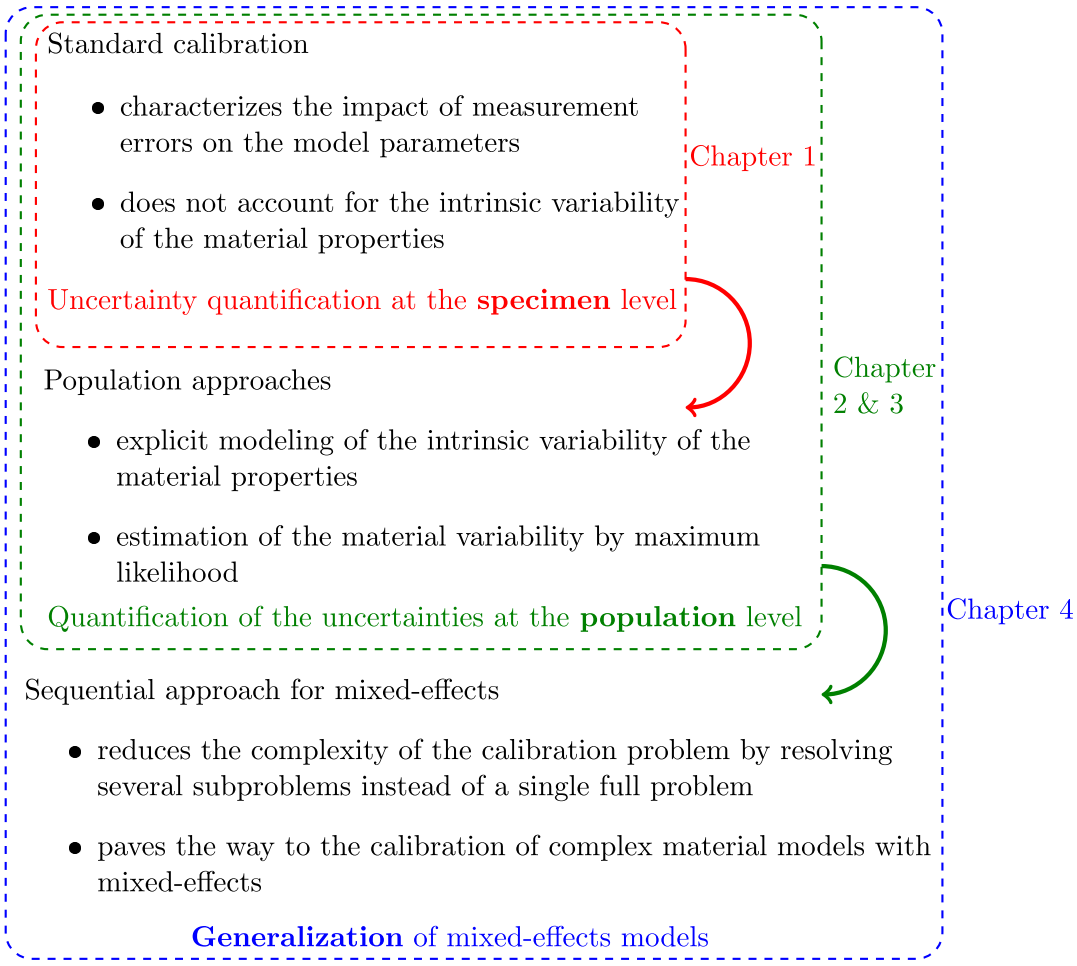


Figure 4.43: Different stages of the work

The main notations and assumptions used throughout the manuscript have been introduced. The techniques to carry out uncertainty quantification in both the frequentist and Bayesian frameworks have been investigated. This corresponds in Figure 4.43 to the red frame, which can be considered as the stage of lowest complexity in our thesis. Yet, an investigation of classical calibration techniques on real experiments has shown that a key assumption of such methods is too restrictive in our application context as it does not account for the impact of material variability (mainly due to the manufacturing process), which turns out to be stronger than the consequences of both measurement errors and model bias. Considering broader frameworks that include material variability explicitly in the statistical modeling of the calibration problem is therefore necessary. In particular, the population approaches, which are already applied in civil engineering, address this problem.

The population approaches, represented by the green rectangle in Figure 4.43, generalize the standard calibration methods. Their main novelty is that they incorporate the modeling of the intrinsic variability of the material properties. Various methods to calibrate this variability can be employed, with some of them applying only in the Bayesian framework (*e.g.*, Hierarchical Bayesian Models). This is a major feature because the Bayesian point of view is in general synonymous with large computational costs. In some situations, a deterministic estimation of the parameters of the joint distribution of the material properties (also called the population parameters) is sufficient. To carry out this estimation, a key method relies on mixed-effects models, on which a focus is made. With mixed-effects, the calibration is achieved by maximizing the likelihood of the observations given the population parameters.

A first implementation of mixed-effects has been carried out to ensure that such methods can be applied to the calibration of mechanical behavior laws. The identification of the individual parameters has been spotted as a key point to ensure a proper calibration of the population parameters. In fact, if standard calibration methods (highlighted in red in Figure 4.43) cannot identify properly the individual parameters on each specimen independently, mixed-effects will not be able to provide an accurate estimation of the population parameters and are not relevant to use. In addition, the identification of the individual parameters is the costly phase of the likelihood computation. The choice of the method implemented to perform this task is particularly sensitive as it has to strike a balance between numerical costs and accuracy of the calibration. For instance, in this manuscript, an evolutionary algorithm (CMA-ES) was replaced with a gradient-based algorithm combined with a multi-start strategy without any loss of accuracy, which enabled to reduce the number of model evaluations. Another complicated point is the number of available independent experiments, i.e., individuals. Of course, in real cases, it is fixed but, with synthetic data, it is possible to evaluate its impact. With few individuals, the population parameters are flawed with significant errors (quantified *e.g.*, with KL-divergence) and uncertainty (quantified through bootstrap) which decreases with the number of individuals. This effect is especially visible on the correlation coefficients that describe the statistical dependencies between the model parameters. Indeed, from our results, with few individuals, the interest of calibrating the statistical dependencies appears to be low. Calibration outside the classical Gaussian framework has also been successfully considered, along with a description of the changes to bring to the estimation of the likelihood function. On the tested example, only minor changes were found to be necessary but the complexity of implementation of the method increased much. In fact, from our observations, a Gaussian multivariate distribution seems to provide the best compromise between complexity and accuracy of the calibration results. The use of non-Gaussian parameter distributions should thus be limited to cases that demonstrate clearly the need to enlarge the modeling hypothesis. These conclusions are nevertheless limited as they concern a specific test case. The proposed method has then be tested on real experiments. As there are no reference values for the population parameters, it was decided to compare the experimental and calibrated material variabilities in model space. This was done by propagating through the model the calibrated distribution with Monte-Carlo techniques. The comparison between the bundle of model outputs from the propagation and the experimental curves showed a satisfactory agreement, in particular when correlations where included. On the contrary, with no statistical dependencies, material variability appeared to be overestimated. This illustrates the need to test several parameters distributions to draw reliable conclusions. Globally, this first study enabled to test the impact of the different settings of the mixed-effects. Another significant point was the inclusion of non-Gaussian marginals that is not often treated. The implementation on the experimental data demonstrated its ability to provide consistent estimates of the observed variability. Nevertheless, the main limit of this work is the simplicity of the model that does not represent the complexity of recent behavior laws.

Composite materials exhibit many different behaviors: elasticity, viscosity, damage, *etc.* This means that the parameters of many models need to be estimated, which makes the estimation of their joint distribution even more complex to perform. In the last chapter, a sequential calibration scheme consistent with the population approaches has been proposed. This proposal widens the classical mixed-effects framework and is represented in blue in Figure 4.43. The different steps of the procedure are defined following an analysis of the behavior law of interest (that refers here to non-linear elasticity). The proposed technique has been implemented using synthetic data and the results compared to the standard application of the mixed-effects method. We have observed in particular that no specific deterioration of the calibration occurred among the different calibration strategies (*e.g.* sequential or standard mixed-effects) and more importantly,

that the sequential scheme allowed a significant gain in terms of model evaluations (divided by 2 in our case).

Perspectives

Although this work has to some extent met the objectives set, it still suffers from several limitations. The first limitation is the modeling of the residuals, that is to say the difference between the experimental data and the calibrated models. Recall that throughout this thesis they have been assumed to be homoscedastic, Gaussian distributed with zero mean and unknown variance. This problem becomes even more significant with real data. Indeed, in the different calibrations carried out, model bias has proven to be more significant than measurement errors. The mixed-effects methodology we have developed is relevant if the probabilistic model can be such that the residuals have a smaller effect than the variability of the material properties. Among the tools that can be used to keep the residual small enough are surrogate models. To sum up, following a first calibration, the structure of the residuals can be learned with a surrogate model, that will be updated throughout the calibration process. It should be underlined that such a procedure would require to update the statistical framework to account for the explicit bias modeling. For example, the likelihood calculation should also be updated to feature the impact of the new model.
Another aspect investigated in this manuscript is the computational costs, which justifies the introduction of a sequential approach. Because of the hierarchical modeling introduced in the population approaches, the computational costs associated with the resolution methods increase, compared to standard calibration methods. A relevant criterion for assessing these costs is the number of model evaluations. In this work, it has been possible to carry out tens of millions of model runs without specific issues, in particular within a reasonable time. However, with more complex models involving the resolution of nonlinear differential equations, this is simply not feasible in a time compatible with design deadlines. A popular answer to address this kind of problems is the use of surrogate models. The first point is to precise at which level such a strategy could be used as it can be implemented either to approximate the likelihood function or the behavior law. If it is decided to build a surrogate model of the likelihood, two errors would accumulate here: the error from the likelihood approximation and the error from the surrogate model. This combination of errors should be cautiously treated, because of the risk of introducing a bias in the population parameters estimates. In particular, this could seriously affect the calibration of the correlations, to which the likelihood function seems less sensitive compared to other parameters. Building a surrogate model of the behavior law also brings several difficulties. Remember that in each of the presented approaches, the likelihood computation requires the estimation of the individual parameters. As the individual parameters can be located in a large space (larger than 20), if such an option is chosen, the surrogate of interest should be valid over a wide range of parameters values. Training such a surrogate model is a significant challenge and may not ensure a satisfactory level of precision making this possibility not fully satisfying. An alternative strategy may be to train the surrogate model locally around the fixed-effects at each iteration. The advantage of such a technique is that the surrogate model may be easier to train, especially if the behavior law has good smoothness properties. However, the estimated individual parameters should be within the range of validity of the trained surrogate model, otherwise, it can mislead the identification of the individual parameters, which would therefore impair the estimation of the population parameters.
Generally, a common limitation of the results provided by the population approaches implemented in this work is related to the number of available independent experiments (individuals). Indeed, with few individuals, it appears that the population parameters can be flawed with high uncertainties. Consequently, the results presented in this manuscript constitute the

first milestone in accounting for specimen variability in constitutive model identification. This research can be continued by quantifying the uncertainties on the population parameters, which can be done in both frequentist and Bayesian inference. In this study, for computational costs reasons, such analyses were not systematically carried out. To consolidate the results, Bayesian inference, in which external knowledge can be added to the calibration problem with the definition of the prior density on the population parameters, seems well-suited. Note that, apart from the difficulties arising from the numerical implementation, the choice of the prior is a major difficulty in the Bayesian paradigm. Of course, a flat but non-informative prior can be chosen but this is not interesting in our context because it does not provide any additional information that could be found either in the data, in the bounds on the population parameters or from expert knowledge. The problem is therefore to design informative priors on the population parameters. To reach this goal, one possibility could consist in exploiting databases of values from former identification on similar materials to infer a prior on the population parameters. Note that the data used to determine the prior should not be used in the calibration process. Another strategy could be based on the multi-scale nature of the composites materials, by propagating uncertainties from the constituents to the ply scale, using an homogenization method.

Among the general perspectives, an interesting axis of development would be to apply the sequential approach in a fixed-point paradigm. Furthermore, the sequential approach has been applied to a case where the results could be compared with the standard application of mixed-effects. To assess the interest of this method, it would be appropriate to apply it to cases where standard mixed-effects calibration is not feasible because of computational costs cumulated with a large number of parameters to estimate. This could provide a better illustration of the efficiency of the developed methods. Thus, the sequential strategy could be implemented to calibrated more complex models than those considered in this thesis. These include complex nonlinear behavior such as viscosity and damage. If mixed-effects enable to characterize the joint distribution of the model parameters for each phenomenon, we can hope to apply the sequential approach to calibrate complete models.
Elementary experiments do not always allow to calibrate all the correlation coefficients between the model parameters. This limit could be overcome with more complex experiments on structures or involving many physics. However, the output data corresponding to these tests are data fields. Data fields, which are much harder to process than either scalars or vectors, usually refer to displacement fields measured with Digital Image Correlation, that is becoming a usual measurements method. The impact on the statistical modeling should be carefully studied. In particular, the data is extracted from images with an inverse method. The data fields are therefore flawed with higher errors than gauges measurements, because the measurement uncertainties combine both measurement noises (that do not originate from the same sources: *e.g.*, lighting, position of the cameras) and the numerical issues from the inverse method. Thus, it should be ensured that the impact of measurements errors remain lower than material variability. In addition, the amount of data to process is much higher, which may slow down the calibration process.
Mixed-effects calibration relies on the identification of the individual parameters. To strengthen their identification, one can think of combining together different types of data (if they are available), *e.g.*, acoustic data and field measurements. This would require a careful review of the statistical formalization but could improve the calibration results.
To increase the number of available individuals, one possibility consists in simultaneously calibrating specimens with different experimental configurations but on which the same parameters can be calibrated. This requires to introduce covariates, parameters that vary from one specimen to the other but that are known. Typically, they refer to the orientations of the plies within a laminate and can be either deterministic or random. Note that in the latter case, the complexity of the likelihood computation increases and deserves to be further investigated.

Finally, to demonstrate the interest of using population approaches, it will be useful, in a certification context, to evaluate the costs of the different methods. On the one hand, the certification would be performed following the usual standards, achieving the recommended certification experiments to estimate the safety factors and the associated costs. On the other hand, calibration with mixed-effects would be performed from experiments repeated at the small scales where they are cheaper to carry out. The estimated variability would then be propagated to higher scale models to predict the behavior of the structures of interest, which can be optimized to find the design that minimizes the manufacturing costs while complying with safety requirements. The performance and the costs associated to each solution could be then compared, such as the corresponding advantages and drawbacks.

APPENDIX A

ESTIMATION OF THE LIKELIHOOD WITH THE LAPLACE APPROXIMATION

The objective of this appendix is to present the practical implementation of the Laplace approximation. Theoritical points are recalled in Section A.1 and the implementation illustrated with a pseudo-code and an explanation scheme in Section A.2.

A.1 Likelihood of mixed-effects

Before introducing the Laplace approximation of the likelihood function, let recall the key notations. The number of available repetitions is noted n with i the corresponding index. The set of model parameters that characterizes the i^{th} specimen (the individual parameters) is labeled $\boldsymbol{\theta}_i, i \in [\![1, n]\!]$. There are d model parameters, so $\boldsymbol{\theta}_i$ is of dimension d. N_i stands for the number of observations of the i^{th} subject. The j^{th} input and output data of the i^{th} specimen are respectively labeled t_{ij} and y_{ij}, with j the index of the j^{th} observation. The behavior law is denoted $\mathcal{F}(\cdot, \cdot)$. The probability distribution that represents material variability and called the model parameters distribution is labeled $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$. Throughout this work, the residuals have been supposed to be Gaussian distributed with a 0 mean and an homoscedastic variance ω_i^2, which is different for all individuals. The full population parameters that combine the population parameters $\boldsymbol{\Pi}$ and those characterizing the residual models, the $[\omega_i^2]_{i\in[\![1,n]\!]}$, are denoted $\boldsymbol{\Psi}$ and refer to $\left[\boldsymbol{\Pi}, \quad \boldsymbol{\omega}\right]$, with $\boldsymbol{\omega} = [\omega_1^2, \ldots, \omega_n^2]$. The dimension of $\boldsymbol{\Psi}$ is denoted p. The population-based approaches seek to determine $\boldsymbol{\Psi}$ and to provide an estimate of the individual parameters $\{\boldsymbol{\theta}_i\}_{i\in[\![1,n]\!]}$ as a by-product. The full population parameters are estimated by maximizing the likelihood of $\boldsymbol{\Psi}$ given the observations $\mathbf{y}_1, \cdots, \mathbf{y}_n$, that is to say $\mathcal{L}(\boldsymbol{\Psi}) = f(\mathbf{y}_1, \cdots, \mathbf{y}_n|\boldsymbol{\Psi})$, with f a generic letter for probability density functions (PDFs). The likelihood reads as [Lavielle, 2014, Pinheiro and Bates, 2002]:

$$\mathcal{L}(\boldsymbol{\Psi}) = \prod_{i=1}^{n} \int_{\mathbb{R}^d} f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) \mathrm{d}\boldsymbol{\theta}_i \ . \tag{A.1}$$

The opposite log-likelihood is approximated with the Laplace approximation as follows [Pinheiro and Bates, 2002, Wolfinger, 1993]:

$$\begin{aligned} -\ln(\mathcal{L}(\boldsymbol{\Psi})) &= \sum_{i=1}^{n} \ln\left(\int_{\mathbb{R}^d} f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) \mathrm{d}\boldsymbol{\theta}_i\right) \\ &\approx \sum_{i=1}^{n} \left(\frac{1}{2}\ln(|\mathcal{H}(g_i)(\widehat{\boldsymbol{\theta}}_i)|) + g_i(\mathbf{y}_i, \widehat{\boldsymbol{\theta}}_i, \boldsymbol{\Psi})\right) - \frac{d}{2}\ln(2\pi) \ , \end{aligned} \tag{A.2}$$

with function $g_i(\cdot)$ defined as:

$$g_i(\mathbf{y}_i, \boldsymbol{\theta}_i, \boldsymbol{\Psi}) \coloneqq -\ln(f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i) f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})) \ . \tag{A.3}$$

In Eq. (A.2), $\widehat{\boldsymbol{\theta}}_i$ refers to the solution of the following minimization problem:

$$\widehat{\boldsymbol{\theta}}_i = \underset{\boldsymbol{\theta}_i \in \Gamma_{\boldsymbol{\theta}}}{\arg\min}\, g_i(\mathbf{y}_i, \boldsymbol{\theta}_i, \boldsymbol{\Psi}) \ , \tag{A.4}$$

with $\Gamma_{\boldsymbol{\theta}}$ the set of admissible parameters representing physical and mathematical constraints. In Eq. (A.2), $\mathcal{H}(g_i)(\cdot)$ is the Hessian of the function $g_i(\cdot)$ defined in Eq. (A.3). It expresses as :

$$\mathcal{H}(g_i)(\widehat{\boldsymbol{\theta}}_i) = -\mathcal{H}(\ln(f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i)))(\widehat{\boldsymbol{\theta}}_i) - \mathcal{H}(\ln(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})))(\widehat{\boldsymbol{\theta}}_i) \ . \tag{A.5}$$

Remember that the residuals are supposed to be Gaussian distributed with a 0 mean and variance $\omega_i^2, i \in [\![1, n]\!]$. Thus, $-\ln(f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i))(\widehat{\boldsymbol{\theta}}_i)$ expresses as $\frac{\|\mathbf{y}_i - \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)\|^2}{2\omega_i^2} + \frac{N_i}{2}\ln(2\pi)$. Its Hessian is approximated as [Bates et al., 1983]:

$$\begin{aligned} -\mathcal{H}(\ln(f(\mathbf{y}_i|\boldsymbol{\theta}_i, \boldsymbol{\Pi}, \omega_i)))(\widehat{\boldsymbol{\theta}}_i) &= \frac{1}{\omega_i^2}\left(\frac{\partial^2 \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_i^\top}(\mathbf{y}_i - \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)) + \frac{\partial \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i}^\top \frac{\partial \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i}\right) \\ &\approx \frac{1}{\omega_i^2}\frac{\partial \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i}^\top \frac{\partial \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i} \end{aligned} \tag{A.6}$$

$\frac{\partial \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i}$ is estimated with finite differences as :

$$\frac{\partial \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i} \approx \frac{\mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i + \delta\widehat{\boldsymbol{\theta}}_i) - \mathcal{F}(\mathbf{t}_i, \widehat{\boldsymbol{\theta}}_i)}{\delta\widehat{\boldsymbol{\theta}}_i} \tag{A.7}$$

The differentiation step, $\delta\widehat{\boldsymbol{\theta}}_i$, is equal to $10^{-3} \times \widehat{\boldsymbol{\theta}}_i$ not to depend on the order of magnitude of the values of the model parameters. The computation of the Hessian of the log-PDF of the model parameters distribution, $\mathcal{H}(\ln(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi})))(\widehat{\boldsymbol{\theta}}_i)$, is studied in the second appendix B. Finally, remember that the model parameters distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ can be uniquely decomposed between some marginals and a copula with the Sklar theorem [Sklar, 1959]:

$$f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) = c(F_1(\theta_{i1}|\boldsymbol{\Pi}_1), \ldots, F_d(\theta_{id}|\boldsymbol{\Pi}_d)|\boldsymbol{\Pi}_{\text{cop}}) \times \prod_{l=1}^{d} f_l(\theta_{il}|\boldsymbol{\Pi}_l) \ , \tag{A.8}$$

with $\boldsymbol{\Pi}_{\text{cop}}$ the copula parameters, $\boldsymbol{\Pi}_l$ the parameters of the l^{th} marginal of distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ (so the marginal marginal distribution of $\theta_{il}, (i, l) \in [\![1, n]\!] \times [\![1, d]\!]$) with probability density function (PDF) $f_l(\cdot)$ and cumulative density function (CDF) $F_l(\cdot)$ and $c(\cdot)$, the density function of the copula. In this manuscript, it is taken as a Gaussian copula, parameterized by a correlation matrix labeled $\mathbf{R}$.

A.2 Numerical procedure for the Laplace approximation of the likelihood of mixed-effects

The computation of the Laplace approximation is described in both Figure A.1 and algorithm 3.

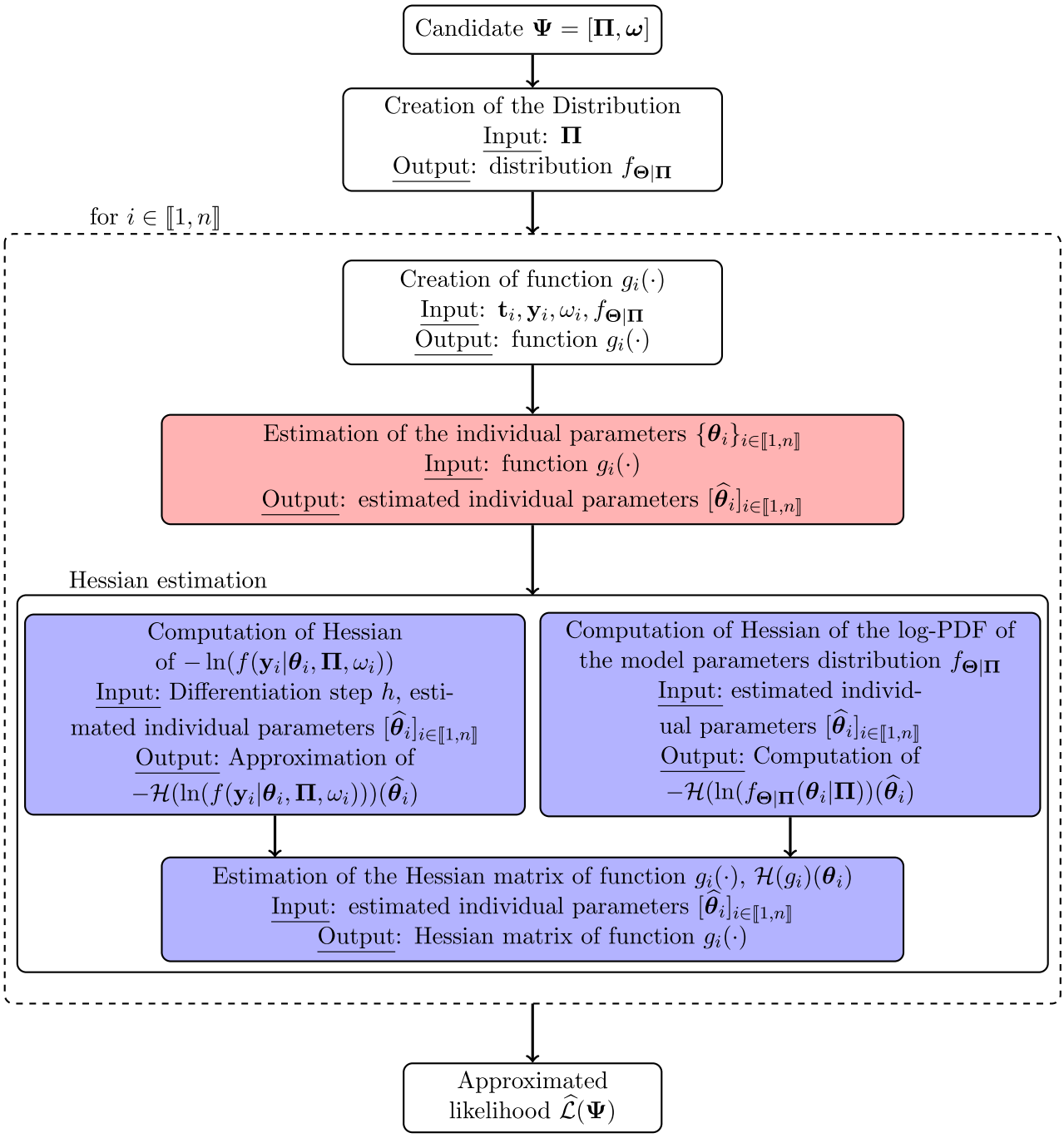


Figure A.1: Computation of the Laplace approximation of the likelihood of mixed-effects. The blue frame refer to the Hessian estimation and those in red to the individual parameters estimation.

Algorithm 3 Algorithm for the computation of the Laplace approximation

Require: Behavior law $\mathcal{F}(\cdot,\cdot)$, experiments $[\mathbf{y}_i]_{i\in[\![1,n]\!]}$, input data, $[\mathbf{t}_i]_{i\in[\![1,n]\!]}$, model parameters distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$

Provide candidate full model parameters $\boldsymbol{\Psi}$ that separates between population parameters $\boldsymbol{\Pi}$ and standard-deviations of the noises $\boldsymbol{\omega}$
if statistical dependencies are included within $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ **then**
    Build correlation matrix ▷ Use spherical decomposition of $\mathbf{R}$ [Pinheiro and Bates, 1996]
end if
Build marginal distributions $l \in [\![1,d]\!]$
Build model parameters distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ ▷ *via* Eq. (A.8)
for $i = 1,\ldots,n$ **do**
    Build function $g_i(\cdot)$ from $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ and PDF $f(\mathbf{y}_i|\boldsymbol{\theta}_i,\boldsymbol{\Pi},\omega_i)$ ▷ *via* Eq. (A.3)
    Estimate the individual parameter $\boldsymbol{\theta}_i$ characterizing the i^{th} specimen by minimizing function $g_i(\cdot)$: $\widehat{\boldsymbol{\theta}}_i = \underset{\boldsymbol{\theta}_i\in\Gamma_{\boldsymbol{\theta}}}{\arg\min}\, g_i(\mathbf{y}_i,\boldsymbol{\theta}_i,\boldsymbol{\Psi})$
    Compute $\frac{\partial\mathcal{F}(\mathbf{t}_i,\widehat{\boldsymbol{\theta}}_i)}{\partial\boldsymbol{\theta}_i}$ by finite differences ▷ *via* Eq. (A.7)
    Computation of the Hessian of the log-PDF of $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ for $\boldsymbol{\theta}_i = \widehat{\boldsymbol{\theta}}_i$ ▷ see details in Appendix B
    Computation of the Hessian of function $g_i(\cdot)$ for $\boldsymbol{\theta}_i = \widehat{\boldsymbol{\theta}}_i$ ▷ *via* Eq. (A.5)
end for
Estimate the log-likelihood $\widehat{\mathcal{L}}$ ▷ *via* Eq. (A.2)

APPENDIX B

DERIVATION OF THE HESSIAN OF THE COPULA DENSITY OF THE MODEL PARAMETERS DISTRIBUTION

The objective of this appendix is to derive the expression of the Hessian of the opposite log-probability density function (PDF) of a Gaussian copula. The first part (Section B.1) aims to recall preliminary results that will be used to compute the Hessian of the quantity of interest in Section B.2. The formulas are then checked on an analytical case in Section B.3, before providing some advice on how to implement them in Section B.4.

B.1 Establishing the formulas

The number of model parameters which corresponds to the dimension of the model parameters is noted d, so the individual parameters $\boldsymbol{\theta}_i$ can be decomposed as $\boldsymbol{\theta}_i = (\theta_{i1}, \dots, \theta_{id})$. Furthermore, remember that the model parameters distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ can be uniquely decomposed between marginals and copula with the Sklar theorem [Sklar, 1959] as follows:

$$f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) = c(F_1(\theta_{i1}|\boldsymbol{\Pi}_1), \dots, F_d(\theta_{id}|\boldsymbol{\Pi}_d)|\boldsymbol{\Pi}_{\text{cop}}) \times \prod_{l=1}^{d} f_l(\theta_{il}|\boldsymbol{\Pi}_l) \ , \tag{B.1}$$

with $\boldsymbol{\Pi}_{\text{cop}}$ the copula parameters, $\boldsymbol{\Pi}_l$ the parameters of the l^{th} marginal of distribution $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}$ (so the distribution of $\theta_{il}, (i,l) \in [\![1,n]\!] \times [\![1,d]\!]$) with probability density function (PDF) $f_l(\cdot)$ and cumulative density function (CDF) $F_l(\cdot)$ and $c(\cdot)$, the density function of the copula. In this manuscript, it is taken as a Gaussian copula, parameterized by a correlation matrix labeled $\mathbf{R}$. Taking the logarithm of Eq. (B.1), the Hessian $\mathcal{H}$ of the log-density writes as

$$\mathcal{H}(\ln(f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}))) = \mathcal{H}(\ln(c(F_1(\theta_{i1}|\boldsymbol{\Pi}_1), \dots, F_d(\theta_{id}|\boldsymbol{\Pi}_d)|\boldsymbol{\Pi}_{\text{cop}})))(\boldsymbol{\theta}_i) + \sum_{l=1}^{d} \mathcal{H}(\ln(f_l(\theta_{il}|\boldsymbol{\Pi}_l)))(\theta_{il}) \ , \tag{B.2}$$

The second term can be computed straightforwardly (standard derivative computation). A focus is made on the calculation of the first term. Remember that here the copula has been chosen as Gaussian, of PDF:

$$c(F_1(\theta_{i1}|\boldsymbol{\Pi}_1), \dots, F_1(\theta_{id}|\boldsymbol{\Pi}_d)|\boldsymbol{\Pi}_{\text{cop}}) \coloneqq \frac{1}{\sqrt{|\mathbf{R}|}} \exp\left(-\frac{1}{2} \begin{pmatrix} \varrho_{i1} \\ \vdots \\ \varrho_{id} \end{pmatrix}^{\top} \mathbf{U} \begin{pmatrix} \varrho_{i1} \\ \vdots \\ \varrho_{id} \end{pmatrix} \right) \ , \tag{B.3}$$

with $\mathbf{R}$ the correlation matrix and $\mathbf{U} \coloneqq \mathbf{R}^{-1} - \mathbf{I}_d$ where $\mathbf{I}_d$ is the identity matrix of $\mathcal{M}_d(\mathbb{R})$. The coefficients of matrix $\mathbf{U}$ are defined as $\mathbf{U} \coloneqq (u_{lm})_{1 \leqslant l,m \leqslant d}$. Variable ϱ_{il} is defined as :

$$\varrho_{il} \coloneqq \Phi^{-1}(F_l(\theta_{il}|\mathbf{\Pi}_l)), l \in [\![1, d]\!] \ , \tag{B.4}$$

with $\Phi^{-1}(\cdot)$ the inverse CDF of univariate Gaussian distribution with a 0 mean and a unit variance. If $\theta_{il} \sim \mathcal{N}(\mu_l, \mathrm{sd}(\theta_{il})^2)$, then $\varrho_{il} = \frac{\theta_{il} - \mu_l}{\mathrm{sd}(\theta_{il})}$. The logarithm of the copula PDF (labeled $\log -c(\cdot|\mathbf{\Pi})$ in the following for the sake of clarity) can be expanded as follows:

$$\begin{aligned} -\log -c(\boldsymbol{\theta}_i|\mathbf{\Pi}) &\coloneqq -\ln(c(F_1(\theta_{i1}|\mathbf{\Pi}_1), \ldots, F_1(\theta_{id}|\mathbf{\Pi}_d)|\mathbf{\Pi}_{\mathrm{cop}})) \\ &= \frac{1}{2}\left(\ln(|\mathbf{R}|) + \sum_{l=1}^{d} u_{ll}\varrho_{il}^2 + 2\sum_{l=1}^{d}\sum_{m=1}^{l-1} u_{lm}\varrho_{il}\varrho_{im}\right) \ . \end{aligned} \tag{B.5}$$

In the following, for the sake of clarity, $F_l(\theta_{i1}|\mathbf{\Pi}_l)$ will be written $F_l(\theta_{il})$ and $f_l(\theta_{il}|\mathbf{\Pi}_l)$ will be replaced with $f_l(\theta_{il})$.

B.2 Derivation of the Hessian

The first derivative of the opposite log-PDF of the copula with respect to θ_{ir} , $r \in [\![1, d]\!]$ is :

$$-\frac{\partial \log -c(\boldsymbol{\theta}_i|\mathbf{\Pi})}{\partial \theta_{ir}} = \frac{1}{2}\left(2u_{rr}\varrho_{ir}\frac{\partial \varrho_{ir}}{\partial \theta_{ir}} + 2\sum_{\substack{m=1 \\ m\neq r}}^{d} u_{rm}\varrho_{im}\frac{\partial \varrho_{ir}}{\partial \theta_{ir}}\right) = \frac{\partial \varrho_{ir}}{\partial \theta_{ir}}\left(u_{rr}\varrho_{ir} + \sum_{\substack{m=1 \\ m\neq r}}^{d} u_{rm}\varrho_{im}\right) \ . \tag{B.6}$$

To terminate the estimate of the first derivative, it remains to compute $\frac{\partial \varrho_{ir}}{\partial \theta_{ir}}$. Remember that for two univariate differentiable real functions $q(\cdot)$ and $w(\cdot)$ that have a reciprocal function with the same regularity properties, we have,

- $(w^{-1})' = \frac{1}{w' \circ w^{-1}}$
- $(w \circ q)' = q' \times (w' \circ q)$ (the chain rule for univariate functions)

with $\circ$ the composition operator. Finally, remember that $\varphi(\cdot)$ is the PDF of the Normal distribution with zero mean and unit variance and $\Phi(\cdot)$ its CDF, so they comply with $\Phi' = \varphi$. The same analysis gives $\forall r \in [\![1, d]\!]$, $F_r' = f_r$. Let apply now the relations recalled above to estimate $\frac{\partial \varrho_{ir}}{\partial \theta_{ir}} = \frac{\partial \Phi^{-1}(F_r(\theta_{ir}))}{\partial \theta_{ir}}$. Let apply the first rule to $w = \Phi^{-1}$. This gives $(\Phi^{-1})' = \frac{1}{\varphi \circ \Phi^{-1}}$. The application of the second formula, taking $w = \Phi^{-1}$ and $q = F_r$, gives $(\Phi^{-1} \circ F_r)' = \frac{f_r}{\varphi \circ \Phi^{-1} \circ F_r}$. Then, for $x = \theta_{ir}, \forall r \in [\![1, d]\!]$, the following result apply:

$$\frac{\partial \varrho_{ir}}{\partial \theta_{ir}} = \frac{f_r(\theta_{ir})}{\varphi(\Phi^{-1}(F_r(\theta_{ir})))} = \frac{f_r(\theta_{ir})}{\varphi(\varrho_{ir})} \ , \tag{B.7}$$

allowing to expand Eq.(B.6) into

$$-\frac{\partial \log -c(\boldsymbol{\theta}_i|\mathbf{\Pi})}{\partial \theta_{ir}} = \frac{f_r(\theta_{ir})}{\varphi(\varrho_{ir})}\left(u_{rr}\varrho_{ir} + \sum_{\substack{m=1 \\ m\neq r}}^{d} u_{rm}\varrho_{im}\right) \ . \tag{B.8}$$

These formulas are analoguous to those obtained in [Schepsmeier and Stöber, 2014] in the case of a bivariate copula. The Hessian of the of the log-PDF of the copula $\log -c$ reads as:

$$\mathcal{H}_{ra}(\log -c) = \frac{\partial^2 \log -c}{\partial \theta_{ir}\partial \theta_{ia}} \ . \tag{B.9}$$

Note that the log-PDF can be supposed without a significant loss of generality twice continuous (with thus a symmetric Hessian matrix). In particular, this enables to invert the order of the derivation. To estimate the Hessian, it remains to derive the second derivative for all indices $(r, a) \in [\![1, d]\!]^2$. Let focus on the simplest case which is when $r \neq a$. Taking the derivative of Eq.(B.8) with respect to θ_{ia} gives:

$$\begin{aligned} -\frac{\partial^2 \log -c(\boldsymbol{\theta}_i|\boldsymbol{\Pi})}{\partial\theta_{ia}\partial\theta_{ir}} &= \frac{\partial}{\partial\theta_{ia}}\left[\frac{\partial \log -c(\boldsymbol{\theta}_i|\boldsymbol{\Pi})}{\partial\theta_{ir}}\right] = \frac{\partial}{\partial\theta_{ia}}\left[\frac{\partial\varrho_{ir}}{\partial\theta_{ir}}\left(u_{rr}\varrho_{ir} + \sum_{\substack{m=1\\m\neq r}}^{d} u_{rm}\varrho_{im}\right)\right] \\ &= \frac{\partial\varrho_{ir}}{\partial\theta_{ir}}\frac{\partial}{\partial\theta_{ia}}\left[\sum_{\substack{m=1\\m\neq r}}^{d} u_{rm}\varrho_{im}\right] = u_{ra}\frac{\partial\varrho_{ir}}{\partial\theta_{ir}}\frac{\partial\varrho_{ia}}{\partial\theta_{ia}} = u_{ra}\frac{f_a(\theta_{ia})}{\varphi(\varrho_{ia})}\frac{f_r(\theta_{ir})}{\varphi(\varrho_{ir})} \ . \end{aligned} \tag{B.10}$$

It remains to estimate the diagonal terms, that is to say when $a = r$. Using the differentiation formulas for sum and product, it gives:

$$\begin{aligned} -\frac{\partial^2 \log -c(\boldsymbol{\theta}_i|\boldsymbol{\Pi})}{\partial\theta_{ir}^2} &= \frac{\partial}{\partial\theta_{ir}}\left[-\frac{\partial \log -c(\boldsymbol{\theta}_i|\boldsymbol{\Pi})}{\partial\theta_{ir}}\right] = \frac{\partial}{\partial\theta_{ir}}\left[\frac{\partial\varrho_{ir}}{\partial\theta_{ir}}\left(u_{rr}\varrho_{ir} + \sum_{\substack{m=1\\m\neq r}}^{d} u_{rm}\varrho_{im}\right)\right] \\ &= \frac{\partial^2\varrho_{ir}}{\partial\theta_{ir}^2}\left(u_{rr}\varrho_{ir} + \sum_{\substack{m=1\\m\neq r}}^{d} u_{rm}\varrho_{im}\right) + u_{rr}\left(\frac{\partial\varrho_{ir}}{\partial\theta_{ir}}\right)^2 \ . \end{aligned} \tag{B.11}$$

To get the expression of the Hessian of the log-PDF of the copula, Eq.(B.10) shows that it is necessary to compute $\frac{\partial^2\varrho_{ir}}{\partial\theta_{ir}^2}$. This gives:

$$\frac{\partial^2\varrho_{ir}}{\partial\theta_{ir}^2} = \frac{\partial}{\partial\theta_{ir}}\left[\frac{\partial\varrho_{ir}}{\partial\theta_{ir}}\right] = \frac{\partial}{\partial\theta_{ir}}\left[\frac{f_r(\theta_{ir})}{\varphi(\varrho_{ir})}\right] = \frac{f_r'(\theta_{ir})\varphi(\varrho_{ir}) - f_r(\theta_{ir})\frac{\partial\varphi(\varrho_{ir})}{\partial\theta_{ir}}}{\varphi(\varrho_{ir})^2} \ . \tag{B.12}$$

The derivative of $\varphi(\varrho_{ir})$ with respect to θ_{ir} (application of the chain rule) expresses as

$$\frac{\partial\varphi(\varrho_{ir})}{\partial\theta_{ir}} = \frac{\partial\varphi(\varrho_{ir})}{\partial\varrho_{ir}}\frac{\partial\varrho_{ir}}{\partial\theta_{ir}} = \varphi'(\varrho_{ir})\frac{f_r(\theta_{ir})}{\varphi(\varrho_{ir})} \ . \tag{B.13}$$

Then, $\frac{\partial^2\varrho_{ir}}{\partial\theta_{ir}^2}$ expresses as:

$$\begin{aligned} \frac{\partial^2\varrho_{ir}}{\partial\theta_{ir}^2} &= \frac{f_r'(\theta_{ir})\varphi(\varrho_{ir}) - f_r(\theta_{ir})\frac{\partial\varphi(\varrho_{ir})}{\partial\theta_{ir}}}{\varphi(\varrho_{ir})^2} = \frac{f_r'(\theta_{ir})\varphi(\varrho_{ir}) - f_r(\theta_{ir})\varphi'(\varrho_{ir})\frac{f_r(\theta_{ir})}{\varphi(\varrho_{ir})}}{\varphi(\varrho_{ir})^2} \\ &= \frac{f_r'(\theta_{ir})\varphi(\varrho_{ir})^2 - f_r(\theta_{ir})^2\varphi'(\varrho_{ir})}{\varphi(\varrho_{ir})^3} \ . \end{aligned} \tag{B.14}$$

Plugging all results together, Eq. (B.11) reduces to:

$$\begin{aligned} -\frac{\partial^2 \log -c(\boldsymbol{\theta}_i|\boldsymbol{\Pi})}{\partial\theta_{ir}^2} &= \frac{\partial^2\varrho_{ir}}{\partial\theta_{ir}^2}\left(u_{rr}\varrho_{ir} + \sum_{\substack{m=1\\m\neq r}}^{d} u_{rm}\varrho_{im}\right) + u_{rr}\left(\frac{\partial\varrho_{ir}}{\partial\theta_{ir}}\right)^2 \\ &= \left[\frac{f_r'(\theta_{ir})\varphi(\varrho_{ir})^2 - f_r(\theta_{ir})^2\varphi'(\varrho_{ir})}{\varphi(\varrho_{ir})^3}\right]\left(u_{rr}\varrho_{ir} + \sum_{\substack{m=1\\m\neq r}}^{d} u_{rm}\varrho_{im}\right) + u_{rr}\left(\frac{f_r(\theta_{ir})}{\varphi(\varrho_{ir})}\right)^2 \ . \end{aligned} \tag{B.15}$$

Finally, the Hessian of the opposite log-PDF of the copula is :

$$-\mathcal{H}_{ra}(\log -c(\boldsymbol{\theta}_i|\boldsymbol{\Pi})) = \begin{cases} \left[\frac{f_r'(\theta_{ir})\varphi(\varrho_{ir})^2 - f_r(\theta_{ir})^2\varphi'(\varrho_{ir})}{\varphi(\varrho_{ir})^3}\right]\left(u_{rr}\varrho_{ir} + \sum_{\substack{m=1\\m\neq r}}^{d} u_{rm}\varrho_{im}\right) + u_{rr}\left(\frac{f_r(\theta_{ir})}{\varphi(\varrho_{ir})}\right)^2 \text{ if } r = a \ , \\ u_{ra}\frac{f_r(\theta_{ir})f_r(\theta_{ir})}{\varphi(\varrho_{ir})\varphi(\varrho_{ir})} \text{ else.} \end{cases} \tag{B.16}$$

B.3 Checking the formulas

The formulas can be checked in the case of normal distributions, that is to say when both the marginals and the copula are Gaussian. Let us present some preliminary results before. Let first remember that a multivariate normal distribution is characterized by a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$. If this is the underlying of the individual parameters, then we have $\boldsymbol{\theta}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with PDF $f_{\boldsymbol{\Theta}|\boldsymbol{\Pi}}(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) = \frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\mu})\right)$. Furthermore, each coordinate θ_{il} is distributed $\mathcal{N}(\mu_l, \mathrm{sd}(\theta_{il})^2)$ The CDF of each marginal $F_l, l \in [\![1, d]\!]$ corresponds to $F_l(\cdot) = \Phi\left(\frac{\cdot - \mu_l}{\mathrm{sd}(\theta_{il})}\right)$. Thus, we have

$$\varrho_{il} \coloneqq \Phi^{-1}(F_l(\theta_{il})) = \Phi^{-1}\left(\Phi\left(\frac{\theta_{il} - \mu_l}{\mathrm{sd}(\theta_{il})}\right)\right) = \frac{\theta_{il} - \mu_l}{\mathrm{sd}(\theta_{il})} \ , \tag{B.17}$$

which gives for the copula PDF

$$\begin{aligned} c(F_1(\theta_{i1}|\boldsymbol{\Pi}_1), \dots, F_1(\theta_{id}|\boldsymbol{\Pi}_d)|\boldsymbol{\Pi}_{\mathrm{cop}}) &\coloneqq \frac{1}{\sqrt{|\mathbf{R}|}} \exp\left(-\frac{1}{2}\begin{pmatrix}\varrho_{i1}\\ \vdots \\ \varrho_{id}\end{pmatrix}^\top \mathbf{U} \begin{pmatrix}\varrho_{i1}\\ \vdots \\ \varrho_{id}\end{pmatrix}\right) \\ &= \frac{1}{\sqrt{|\mathbf{R}|}} \exp\left(-\frac{1}{2}(\mathbf{S}^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\mu}))^\top \mathbf{U} (\mathbf{S}^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\mu}))\right) \ , \end{aligned} \tag{B.18}$$

with $\mathbf{S}$ a diagonal matrix whose terms correspond to the standard-deviations of each coordinate. Its log-PDF simply reduces to a quadratic form :

$$\log -c(\boldsymbol{\theta}_i|\boldsymbol{\Pi}) = -\frac{1}{2}\left(\ln(|\mathbf{R}|) + \sum_{l=1}^d u_{ll}\left(\frac{\theta_{il}-\mu_l}{\mathrm{sd}(\theta_{il})}\right)^2 + 2\sum_{l=1}^d\sum_{m<l} u_{lm}\left(\frac{\theta_{il}-\mu_l}{\mathrm{sd}(\theta_{il})}\right)\left(\frac{\theta_{im}-\mu_m}{\mathrm{sd}(\theta_{im})}\right)\right) \ , \tag{B.19}$$

whose Hessian matrix is easy to derive and refers to:

$$-\mathcal{H}_{ra}(\log -c(\boldsymbol{\theta}_i|\boldsymbol{\Pi})) = \begin{cases} \frac{u_{rr}}{\mathrm{sd}(\theta_{ir})^2} \text{ if } r = a \ , \\ \\ \frac{u_{ra}}{\mathrm{sd}(\theta_{ir})\mathrm{sd}(\theta_{ia})} \text{ else.} \end{cases} \tag{B.20}$$

Before checking the formulas, let notice that for a Gaussian distributed parameter, that is to say $\theta_{ir} \sim \mathcal{N}(\mu_r, \mathrm{sd}(\theta_{ir})^2)$, its PDF, $f_r(\cdot)$, complies with $f_r(\theta_{ir}) = \frac{1}{\mathrm{sd}(\theta_{ir})}\varphi\left(\frac{\theta_{ir}-\mu_r}{\mathrm{sd}(\theta_{ir})}\right)$ and its derivative verifies $f_r'(\theta_{ir}) = \frac{1}{\mathrm{sd}(\theta_{ir})^2}\varphi'\left(\frac{\theta_{ir}-\mu_r}{\mathrm{sd}(\theta_{ir})}\right)$ (following the chain rule).

The first point to be checked is the second derivative of ϱ_{ir} with respect to θ_{ir}, is zero in this specific case following Eq.(B.17). If we take Eq. (B.14), we have

$$\begin{aligned} \frac{\partial^2 \varrho_{ir}}{\partial \theta_{ir}^2} &= \frac{f_r'(\theta_{ir})\varphi(\varrho_{ir})^2 - f_r(\theta_{ir})^2\varphi'(\varrho_{ir})}{\varphi(\varrho_{ir})^3} \\ &= \frac{\frac{1}{\mathrm{sd}(\theta_{ir})^2}\varphi'\left(\frac{\theta_{ir}-\mu_r}{\mathrm{sd}(\theta_{ir})}\right)\varphi\left(\frac{\theta_{ir}-\mu_r}{\mathrm{sd}(\theta_{ir})}\right)^2 - \left(\frac{1}{\mathrm{sd}(\theta_{ir})}\varphi\left(\frac{\theta_{ir}-\mu_r}{\mathrm{sd}(\theta_{ir})}\right)\right)^2\varphi'\left(\frac{\theta_{ir}-\mu_r}{\mathrm{sd}(\theta_{ir})}\right)}{\varphi\left(\frac{\theta_{ir}-\mu_r}{\mathrm{sd}(\theta_{ir})}\right)^3} \\ &= \frac{\varphi'\left(\frac{\theta_{ir}-\mu_r}{\mathrm{sd}(\theta_{ir})}\right)\varphi\left(\frac{\theta_{ir}-\mu_r}{\mathrm{sd}(\theta_{ir})}\right)^2 - \varphi\left(\frac{\theta_{ir}-\mu_r}{\mathrm{sd}(\theta_{ir})}\right)^2\varphi'\left(\frac{\theta_{ir}-\mu_r}{\mathrm{sd}(\theta_{ir})}\right)}{\mathrm{sd}(\theta_{ir})^2\varphi\left(\frac{\theta_{ir}-\mu_r}{\mathrm{sd}(\theta_{ir})}\right)^3} = 0 \ , \end{aligned} \tag{B.21}$$

which leaves, for the diagonal term of the Hessian matrix, using the relation between f_r and φ recalled above,

$$
\begin{aligned}
-\frac{\partial^2 \log -c(\boldsymbol{\theta}_i|\boldsymbol{\Pi})}{\partial \theta_{ir}^2} &= \frac{\partial^2 \varrho_{ir}}{\partial \theta_{ir}^2}\left(u_{rr}\varrho_{ir} + \sum_{\substack{m=1\\m\neq r}}^{d} u_{rm}\varrho_{im}\right) + u_{rr}\left(\frac{\partial \varrho_{ir}}{\partial \theta_{ir}}\right)^2 \\
&= \left[\frac{f_l'(\theta_{ir})\varphi(\varrho_{ir})^2 - f_r(\theta_{ir})^2\varphi'(\varrho_{ir})}{\varphi(\varrho_{ir})^3}\right]\left(u_{rr}\varrho_{ir} + \sum_{\substack{m=1\\m\neq r}}^{d} u_{rm}\varrho_{im}\right) + u_{rr}\left(\frac{f_r(\theta_{ir})}{\varphi(\varrho_{ir})}\right)^2 \\
&= 0 \times \left(u_{rr}\varrho_{ir} + \sum_{\substack{m=1\\m\neq r}}^{d} u_{rm}\varrho_{im}\right) + u_{rr}\left(\frac{f_r(\theta_{ir})}{\varphi(\varrho_{ir})}\right)^2 = u_{rr}\left(\frac{1}{\mathrm{sd}(\theta_{ir})}\right)^2 = \frac{u_{rr}}{\mathrm{sd}(\theta_{ir})^2}\,,
\end{aligned}
\tag{B.22}
$$

which is the expected result according to Eq. (B.20). The cross derivates can also checked:

$$
-\frac{\partial^2 \log -c(\boldsymbol{\theta}_i|\boldsymbol{\Pi})}{\partial \theta_{ir}\partial \theta_{ia}} = u_{ra}\frac{f_r(\theta_{ir})f_a(\theta_{ia})}{\varphi(\varrho_{ir})\varphi(\varrho_{ia})} = u_{ra} \times \frac{1}{\mathrm{sd}(\theta_{ir})} \times \frac{1}{\mathrm{sd}(\theta_{ir})} = \frac{u_{ra}}{\mathrm{sd}(\theta_{ia})\mathrm{sd}(\theta_{ia})}\,. \tag{B.23}
$$

In Eq. (B.23), the links between f_r, f_a and φ have been used. It provides the same result as presented in Eq. (B.20), asserting that the exactness of the formulas in the Gaussian case.

B.4 Practical recommendations for implementation

We have supposed that the log-PDF of the copula was twice continuous with respect to $\boldsymbol{\theta}_i$. Thus, its Hessian matrix is symmetric. Consequently, in the numerical implementation, only the upper(lower) part of the matrix should be estimated. In addition, notice that the mixed derivatives express somehow easily with respect to the ratio $\frac{f_l(\theta_{ir})}{\varphi(\varrho_{ir})}$ and $\frac{f_a(\theta_{ia})}{\varphi(\varrho_{ia})}$. The extra-diagonal terms can be therefore estimated with appropriate vector operations. The diagonal terms are more difficult to handle and should processed carefully.

To ensure the highest level of accuracy, it is advised to discriminate between the Gaussian and non-Gaussian variables. Indeed, as shown in Eq. (B.20), the terms of the Hessian matrix for the Gaussian distributed parameters are easier to compute and more precise. In particular, the complex part of the formula, that is to say $\Phi^{-1} \circ F_l$, is not present. It is also true for the diagonal terms that express as a simple ratio instead of the complex expressions detailed in Eq. (B.16). It should be noted that, from a numerical point of view, the greatest weakness of the proposed formulas is the estimation of $\varphi(\varrho_{ir})$, especially when the individual parameters are far from the average. Indeed, if θ_{ir} is far from the mean(mode) of the distribution, (for instance $|\theta_{ir} - \mu_r| > 7\mathrm{sd}(\theta_{ir})$ for a Gaussian distributed variable), then, $F_r(\theta_{ir}) \to 1$ (or equivalently to zero), which implies that $\varrho_{ir} = \Phi^{-1}(F_l(\theta_{ir})) \to \pm\infty$ and consequently, $\varphi(\varrho_{ir}) \to 0$. Provided that this term appears at the denominator of all terms of the Hessian matrix following Eq. (B.16), this point should not be forgotten. Thresholds can solve to some extent this issue.

BIBLIOGRAPHY

[SAS, 2004] (2004). Sas. 9.1.2. SAS Institute Inc., ed. Cary.

[nml, 2008] (2008). Matlab module nlmefit. https://fr.mathworks.com/help/stats/nlmefit.html.

[Aarons, 1993] Aarons, L. (1993). The estimation of population pharmacokinetic parameters using an em algorithm. *Computer Methods and Programs in Biomedicine*, 41(1):9–16.

[Akaike, 1974] Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.

[Allen, 1974] Allen, D. (1974). The relationship between variable selection and data agumentation and a method for prediction. *Technometrics*, 16(1):125–127.

[Almquist et al., 2015] Almquist, J., Leander, J., and Jirstrand, M. (2015). Using sensitivity equations for computing gradients of the foce and focei approximations to the population likelihood. *Journal of pharmacokinetics and pharmacodynamics*, 42(3):191–209.

[An et al., 2021] An, H., Youn, B., and Kim, H. (2021). Reliability-based design optimization of laminated composite structures under delamination and material property uncertainties. *International Journal of Mechanical Sciences*, 205:106561.

[Anane et al., 2019] Anane, E., López C, D., Barz, T., Sin, G., Gernaey, K., Neubauer, P., and Cruz Bournazou, M. (2019). Output uncertainty of dynamic growth models: Effect of uncertain parameter estimates on model reliability. *Biochemical Engineering Journal*, 150:107247.

[Anderson and Burnham, 2004] Anderson, D. and Burnham, K. (2004). *Model selection and multi-model inference*. Springer New York, NY.

[Andrieu et al., 2003] Andrieu, C., De Freitas, N., Doucet, A., and Jordan, M. (2003). An introduction to MCMC for machine learning. *Machine learning*, 50(01):5–43.

[Antoniadis and Carmona, 1992] Antoniadis, A. and Carmona, R. (1992). *Régression non linéaire et applications (in french)*. Economie et statistiques avançées. Economica.

[Arendt et al., 2012] Arendt, P., Apley, D., and Chen, W. (2012). Quantification of model uncertainty: Calibration, model discrepancy, and identifiability. *Journal of Mechanical Design*, 134:100908.

[Ashlock, 2006] Ashlock, D. (2006). *Evolutionary computation for modeling and optimization*, volume 571. Springer.

[Avril et al., 2008] Avril, S., Bonnet, S., Bretelle, A.-S., Grédiac, M., Hild, F., Ienny, P., Latourte, F., Lemosse, S., Pagano, S., Pagnacco, E., and Pierron, F. (2008). Overview of identification methods of mechanical parameters based on full-field measurements. *Experimental Mechanics*, 48:381–402.

[Baey et al., 2016] Baey, C., Trevezas, S., and Cournède, P.-H. (2016). A nonlinear mixed-effects model of plant growth and estimation via stochastic variants of the em algorithm. *Communications in Statistics - Theory and Methods*, 45(6):1643–1669.

[Ballesteros et al., 2014] Ballesteros, G., Angelikopoulos, P., Papadimitriou, C., and Koumoutsakos, P. (2014). Bayesian hierarchical models for uncertainty quantification in structural dynamics. *Vulnerability, uncertainty, and risk: Quantification, mitigation, and management*, 162:1615–1624.

[Banks et al., 2014] Banks, H. T., Hu, S., and Thompson, W. (2014). *Modeling and Inverse Problems in the Presence of Uncertainty*. Chapman and Hall/CRC.

[Barata and Hussein, 2011] Barata, J. and Hussein, M. (2011). The moore–penrose pseudoinverse: a tutorial review of the theory. *Brazilian Journal of Physics*, 42:146–165.

[Barndorff-Nielsen, 1983] Barndorff-Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, 70(2):343–365.

[Bates et al., 1983] Bates, D., Hamilton, D., and Watts, D. (1983). Calculation of intrinsic and parameter-effects curvatures for nonlinear regression models. *Communications in Statistics - Simulation and Computation*, 12(4):469–477.

[Baudin et al., 2016] Baudin, M., Dutfoy, A., Iooss, B., and Popelin, A.-L. (2016). OpenTURNS: An industrial software for uncertainty quantification in simulation. In Ghanem, R., Higdon, D., and Owhadi, H., editors, *Handbook of Uncertainty Quantification*, pages 1–38. Springer International Publishing.

[Bauer, 2019] Bauer, R. (2019). Nonmem tutorial part i: Description of commands and options, with simple examples of population analysis. *CPT: Pharmacometrics & Systems Pharmacology*, 8(8):525–537.

[Beal and Sheiner, 1982] Beal, S. and Sheiner, L. (1982). Estimating population kinetics. *Critical reviews in biomedical engineering*, 8(3):195—222.

[Beck and de Santana Gomes, 2012] Beck, A. and de Santana Gomes, W. (2012). A comparison of deterministic, reliability-based and risk-based structural optimization under uncertainty. *Probabilistic Engineering Mechanics*, 28:18–29.

[Behmanesh et al., 2015] Behmanesh, I., Moaveni, B., Lombaert, G., and Papadimitriou, C. (2015). Hierarchical bayesian model updating for structural identification. *Mechanical Systems and Signal Processing*, 64-65:360–376.

[Ben Ramdane, 2014] Ben Ramdane, C. (2014). *Etude et modélisation du comportement mécanique de CMC oxyde/oxyde (in french)*. Ph.d. thesis, Université de Bordeaux.

[Ben Ramdane et al., 2017] Ben Ramdane, C., Julian-Jankowiak, A., Valle, R., Renollet, Y., Parlier, M., Martin, E., and Diss, P. (2017). Microstructure and mechanical behaviour of a nextel™610/alumina weak matrix composite subjected to tensile and compressive loadings. *Journal of the European Ceramic Society*, 37(8):2919–2932.

[Berger, 1985] Berger, J. (1985). *Statistical decision theory and Bayesian analysis; 2nd ed.* Springer Science & Business Media.

[Berthelot, 2012] Berthelot, J.-M. (2012). *Matériaux Composites—Comportement Mécanique et Analyse des Structures (in french)*. Lavoisier.

[Bouloré, 2019] Bouloré, A. (2019). Importance of uncertainty quantification in nuclear fuel behaviour modelling and simulation. *Nuclear Engineering and Design*, 355:110311.

[Box and Tiao, 2011] Box, G. and Tiao, G. (2011). *Bayesian inference in statistical analysis*. John Wiley & Sons.

[Bright et al., 2011a] Bright, G., Kennedy, J., Robinson, F., Evans, M., Whittaker, M., Sullivan, J., and Gao, Y. (2011a). Variability in the mechanical properties and processing conditions of a high strength low alloy steel. *Procedia Engineering*, 10:106–111.

[Bright et al., 2011b] Bright, G., Kennedy, J., Robinson, F., Evans, M., Whittaker, M., Sullivan, J., and Gao, Y. (2011b). Variability in the mechanical properties and processing conditions of a high strength low alloy steel. *Procedia Engineering*, 10:106–111.

[Camanho et al., 2006] Camanho, P., Dávila, C., Pinho, S., Iannucci, L., and Robinson, P. (2006). Prediction of in-situ strengths and matrix cracking in composites under transverse tension and in-plane shear. *Composites Part A: Applied Science and Manufacturing*, 37(2):165–176.

[Carpenter and Bithell, 2000] Carpenter, J. and Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. *Statistics in Medicine*, 19(9):1141–1164.

[Carroll and Ruppert, 1988] Carroll, R. and Ruppert, D. (1988). *Transformation and Weighting in Regression*. Springer US.

[Chambers et al., 2018] Chambers, J., Cleveland, W., Kleiner, B., and Tukey, P. (2018). *Graphical methods for data analysis*. Chapman and Hall/CRC.

[Chen et al., 2017] Chen, L., Fan, Y., and Chen, L. (2017). Global sensitivity analysis for fiber reinforced composite fiber path based on d-morph-hdmr algorithm. *Structural and Multidisciplinary Optimization*, 56:1–16.

[Chongshuai et al., 2019] Chongshuai, W., Yiqian, H., and Haitian, Y. (2019). A SBFEM and sensitivity analysis based algorithm for solving inverse viscoelastic problems. *Engineering Analysis with Boundary Elements*, 106:588–598.

[Cintrón-Arias et al., 2009] Cintrón-Arias, A., Banks, H., Capaldi, A., and Lloyd, A. (2009). A sensitivity matrix based methodology for inverse problem formulation. 17(6):545–564.

[Cobelli and DiStefano, 1980] Cobelli, C. and DiStefano, J. (1980). Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 239.

[Coello, 2000] Coello, C. (2000). An updated survey of ga-based multi-objective optimization techniques. *Association for Computing Machinery Computing Surveys*, 32(2):109–143.

[Coello, 2005] Coello, C. (2005). *Recent Trends in Evolutionary Multiobjective Optimization*. Springer London.

[Cokelaer, 2008] Cokelaer, T. (2008). Parameter estimation of inspiralling compact binaries in ground-based detectors: comparison between monte carlo simulations and the fisher information matrix. *Classical and Quantum Gravity*, 25(18):184007.

[Colby and Bair, 2013] Colby, E. and Bair, E. (2013). Cross-validation for nonlinear mixed-effects models. *Journal of pharmacokinetics and pharmacodynamics*, 40:243–252.

[Collette and Siarry, 2002] Collette, Y. and Siarry, P. (2002). *Optimisation Multiobjectif (in french)*. Eyrolles.

[Congdon, 2007] Congdon, P. (2007). *Bayesian statistical modelling*. John Wiley & Sons.

[Cornillon and Matzner-Løber, 2007] Cornillon, P.-A. and Matzner-Løber, E. (2007). *Régression: théorie et applications (in french)*. Springer.

[Cover and Thomas, 2006] Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience.

[Currey et al., 2007] Currey, J., Pitchford, J., and Baxter, P. (2007). Variability of the mechanical properties of bone, and its evolutionary consequences. *Journal of the Royal Society Interface*, 4(12):127–135.

[Damlen et al., 1999] Damlen, P., Wakefield, J., and Walker, S. (1999). Gibbs sampling for bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2):331–344.

[Daniel and Ishai, 2006] Daniel, I. and Ishai, O. (2006). *Engineering Mechanics of Composite Materials*. Oxford university press New York.

[Das and Krishen, 1999] Das, S. and Krishen, A. (1999). Some bootstrap methods in nonlinear mixed-effect models. *Journal of Statistical Planning and Inference*, 75(2):237–245.

[Davidian and Giltinan, 1995] Davidian, M. and Giltinan, D. (1995). *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

[Defense et al., 1999] Defense, U., Company, T. P., Corporation, M. S., for Testing, A. S., and Materials (1999). *Composite Materials Handbook-MIL 17, Volume III: Materials Usage, Design, and Analysis*. The Composite Materials Handbook-MIL 17. Taylor & Francis.

[Delattre et al., 2015] Delattre, M., Lavielle, M., and Poursat, M.-A. (2015). A note on bic in mixed-effects models. *Electronic Journal of Statistics*, 8:456–475.

[Delyon et al., 1999] Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a stochastic approximation version of em algorithm. *The Annals of Statistics*, 27:94–128.

[Demidenko, 2013] Demidenko, E. (2013). *Mixed models. Theory and applications with R.* John Wiley & Sons.

[Dempster et al., 1977] Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.

[Devroye, 1986] Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-New York.

[Drikvandi, 2017] Drikvandi, R. (2017). Nonlinear mixed-effects models for pharmacokinetic data analysis: assessment of the random-effects distribution. *Journal of Pharmacokinetics and Pharmacodynamics*, 44(3):223–232.

[Débarre et al., 2022] Débarre, A., Julian-Jankowiak, A., Parlier, M., Renollet, Y., Pujol, G., and Boussuge, M. (2022). Effect of temperature on the mechanical behaviour of an oxide/oxide composite. *Journal of the European Ceramic Society*, 42(15):7149–7156.

[Economides et al., 2021] Economides, A., Arampatzis, G., Alexeev, D., Litvinov, S., Amoudruz, L., Kulakova, L., Papadimitriou, C., and Koumoutsakos, P. (2021). Hierarchical bayesian uncertainty quantification for a model of the red blood cell. *Physical Review Applied*, 15(3):034062.

[Efron, 1979] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

[Efron and Tibshirani, 1994] Efron, B. and Tibshirani, R. (1994). *An introduction to bootstrap*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman & Hall, 1 edition.

[Efstratiadis and Koutsoyiannis, 2010] Efstratiadis, A. and Koutsoyiannis, D. (2010). One decade of multi-objective calibration approaches in hydrological modelling: a review. *Hydrological Sciences Journal*, 55(1):58–78.

[Ehrgott, 1999] Ehrgott, M. (1999). A characterization of lexicographic max-ordering solutions. (12).

[EPA, 2021] EPA (2021). Control of air pollution from airplanes and airplane engines: Ghg emission standards and test procedures. https://www.epa.gov/regulations-emissions-vehicles-and-engines/control-air-pollution-airplanes-and-airplane-engines-ghg.

[Ette and Williams, 2004] Ette, E. and Williams, P. (2004). Population pharmacokinetics i: Background, concepts, and models. *The Annals of pharmacotherapy*, 38(10):1702–1706.

[Everitt and Skrondal, 2010] Everitt, B. and Skrondal, A. (2010). *The Cambridge Dictionary of Statistics*. Cambridge University Press.

[Fernández-Martínez et al., 2011] Fernández-Martínez, J., Mukerji, T., García-Gonzalo, E., and Fernández-Muñiz, Z. (2011). Uncertainty assessment for inverse problems in high dimensional spaces using particle swarm optimization and model reduction techniques. *Mathematical and Computer Modelling*, 54(11-12):2889–2899.

[Fidler et al., 2019] Fidler, M., Wilkins, J., Hooijmaijers, R., Post, T., Schoemaker, R., Trame, M., Xiong, Y., and Wang, W. (2019). Nonlinear mixed-effects model development and simulation using nlmixr and related r open-source packages. *CPT: Pharmacometrics & Systems Pharmacology*, 8(9):621–633.

[Fisher, 1919] Fisher, R. (1919). The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52:399–438.

[Foreman-Mackey et al., 2013] Foreman-Mackey, D., Hogg, D., Lang, D., and Goodman, J. (2013). emcee: The MCMC hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306–312.

[Forrester et al., 2008] Forrester, A., Sobester, A., and Keane, A. (2008). *Engineering Design via Surrogate Modelling - A Practical Guide*. John Wiley & Sons.

[Fougerouse, 2023] Fougerouse, C. (2023). *Understanding and modelling of the effects of out-of-plane waviness defects on the mechanical performance of a thermoplastic matrix laminate*. Ph.d. thesis, Université Paris-Saclay.

[Frazier, 2018] Frazier, P. I. (2018). A tutorial on bayesian optimization.

[Fu et al., 2013] Fu, L., Wang, M., Sharma, R., and Tang, S. (2013). Parameter estimation of nonlinear mixed-effects models using first order conditional linearization and the em algorithm. *Journal of Applied Statistics*, 40(2):252–265.

[Gallagher and Doherty, 2007] Gallagher, M. and Doherty, J. (2007). Parameter estimation and uncertainty analysis for a watershed model. *Environmental Modelling & Software*, 22(7):1000–1020.

[Gelman et al., 2014] Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2014). *Bayesian data analysis*. Chapman & Hall/CRC texts in statistical science. Chapman and Hall/CRC, third edition. edition.

[Gelman et al., 1996] Gelman, A., Roberts, G. O., and Gilks, W. R. (1996). *Efficient Metropolis jumping rules*, pages 599–608. Oxford University Press, Oxford.

[Germain, 2020] Germain, J. (2020). *Évaluation des capacités prédictives d'un modèle avancé pour la prévision de la tenue de plaques stratifiées perforées (in french)*. Ph.d. thesis, Université Paris-Saclay, Onera.

[Geweke, 1989] Geweke, J. (1989). Bayesian inference in econometric models using monte carlo integration. *Econometrica*, 57(6):1317–1339.

[Ghanem et al., 1991] Ghanem, R., Ghanem, R., and Spanos, P. (1991). *Stochastic Finite Elements: A Spectral Approach*.

[Gibanica et al., 2016] Gibanica, M., Abrahamsson, T., and Olsson, M. (2016). *Calibration, Validation and Uncertainty Quantification of Nominally Identical Car Subframes*, pages 315–326. Springer International Publishing.

[Gilbert, 1991] Gilbert, G. (1991). Positive definite matrices and sylvester's criterion. *The American Mathematical Monthly*, 98(1):44–46.

[Gill, 2007] Gill, J. (2007). Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences. Taylor & Francis.

[Girden, 1992] Girden, E. R. (1992). *ANOVA: Repeated measures*. Number 84. SAGE.

[Gogu, 2009] Gogu, C. (2009). *Faciliter l'identification bayesienne des propriétés élastiques par réduction de dimensionnalité et la méthode des surfaces de réponse*. Ph.d. thesis, ED-488 SIS.

[Gogu et al., 2013] Gogu, C., Yin, W., Haftka, R., Ifju, P., Molimard, J., Le Riche, R., and Vautrin, A. (2013). Bayesian identification of elastic constants in multi-directional laminate from moiré interferometry displacement fields. *Experimental Mechanics*, 53(4):635–648.

[Grail, 2013] Grail, G. (2013). *Approche multimodèle pour la conception de structures composites à renfort tissé (in french)*. Ph.d. thesis, Université d'Orléans.

[Greven and Kneib, 2010] Greven, S. and Kneib, T. (2010). On the behaviour of marginal and conditional aic in linear mixed models. *Biometrika*, 97(4):773–789.

[Guedj et al., 2007] Guedj, J., Thiébaut, R., and Commenges, D. (2007). Maximum likelihood estimation in dynamical models of hiv. *Biometrics*, 63(4):1198–1206.

[Guida and Pulcini, 2009] Guida, M. and Pulcini, G. (2009). Reliability analysis of mechanical systems with bounded and bathtub shaped intensity function. *IEEE Transactions on Reliability*, 58(3):432–443.

[Gusev et al., 2000] Gusev, A., Hine, P., and Ward, I. (2000). Fiber packing and elastic properties of a transversely random unidirectional glass/epoxy composite. *Composites Science and Technology*, 60(4):535–541.

[Haario et al., 1999] Haario, H., Saksman, E., and Tamminen, J. (1999). Adaptive proposal distribution for random walk metropolis algorithm. 14:375–395.

[Hadka, 2015] Hadka, D. (2015). Platypus module. https://platypus.readthedocs.io/en/latest/index.html.

[Hahn and Tsai, 1973] Hahn, H. and Tsai, S. (1973). Nonlinear elastic behavior of unidirectional composite laminae. *Journal of Composite Materials*, 7(1):102–118.

[Hale, 2006] Hale, J. (2006). 787 from the ground up. *Aero Magazine*, 4.

[Hall and Clutter, 2004] Hall, D. and Clutter, M. (2004). Multivariate multilevel nonlinear mixed-effects models for timber yield predictions. *Biometrics*, 60(1):16–24.

[Hansen et al., 2019] Hansen, N., Akimoto, Y., and Baudis, P. (2019). CMA-ES/pycma on Github.

[Hansen et al., 2003] Hansen, N., Müller, S., and Koumoutsakos, P. (2003). Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary computation*, 11(1):1–18.

[Hansen and Ostermeier, 2001] Hansen, N. and Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195.

[Hansen et al., 1997] Hansen, N., Ostermeier, A., and Gawelczyk, A. (1997). On the adaptation of arbitrary normal mutation distributions in evolution strategies: The generating set adaptation. *Proceedings of the Sixth International Conference on Genetic Algorithms*, pages 57–64.

[Harris et al., 2020] Harris, C., Millman, K. J., van der Walt, S., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M., Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. (2020). Array programming with NumPy. *Nature*, 585:357–362.

[Harris and Lee, 2014] Harris, S. and Lee, W. (2014). A study of decline curve analysis in the elm coulee field. In *SPE Unconventional Resources Conference*.

[Hashin, 1980] Hashin, Z. (1980). Failure criteria for unidirectional fiber composites. *Journal of Applied Mechanics*, 47(2):329–334.

[Hastings, 1970] Hastings, W. (1970). Monte-carlo sampling methods using markov chains and their applications. *Biometrika*, 57:14.

[He et al., 2023] He, C., Zhang, Y., Gong, D., and Ji, X. (2023). A review of surrogate-assisted evolutionary algorithms for expensive optimization problems. *Expert Systems with Applications*, 217:119495.

[Heiermann et al., 2005] Heiermann, K., Riesch-Oppermann, H., and Huber, N. (2005). Reliability confidence intervals for ceramic components as obtained from bootstrap methods and neural networks. *Computational Materials Science*, 34(1):1–13.

[Henderson et al., 1959] Henderson, C., Kempthorne, O., Searle, S., and Von Krosigk, C. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, 15(2):192–218.

[Herbert Robbins and Monro, 1951] Herbert Robbins, H. and Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400 – 407.

[Hexcel, 2020] Hexcel (2020). Hexply®m21 global datasheet. https://www.hexcel.com/user_area/content_media/raw/HexPly_M21_global_DataSheet.pdf.

[Higgins et al., 1997] Higgins, K., Davidian, M., and Giltinan, D. (1997). A two-step approach to measurement error in time-dependent covariates in nonlinear mixed-effects models, with application to igf-i pharmacokinetics. *Journal of the American Statistical association*, 92(438):436–448.

[Horowitz, 1996] Horowitz, J. (1996). *Bootstrap methods in econometrics: theory and numerical performance*. Cambridge University press.

[Hu et al., 2017] Hu, Z., Ao, D., and Mahadevan, S. (2017). Calibration experimental design considering field response and model uncertainty. *Computer Methods in Applied Mechanics and Engineering*, 318:92–119.

[Hu et al., 2011] Hu, Z., Smith, R., Hays, M., and Oates, W. (2011). Statistical parameter estimation and uncertainty quantification for macro fiber composite actuators operating in nonlinear and hysteretic regimes. In *50th IEEE Conference on Decision and Control and European Control Conference*, volume 0, pages 2764–2769.

[Hyslop and Imbens, 2001] Hyslop, R. and Imbens, G. (2001). Bias from classical and other forms of measurement error. *Journal of Business & Economic Statistics*, 19(4):475–481.

[Jacquet et al., 2017] Jacquet, E., Chambert, J., Pauchot, J., and Sandoz, P. (2017). Intra- and inter-individual variability in the mechanical properties of the human skin from in vivo measurements on 20 volunteers. *Skin Research and Technology*, 23(4):491–499.

[Jahan and Edwards, 2013] Jahan, A. and Edwards, K. (2013). *Multi-Criteria Decision Analysis for Supporting the Selection of Engineering Materials in Product Design*.

[Jain et al., 2005] Jain, A., Jianchang, M., and Mohiuddin, K. (2005). Artificial neural networks: a tutorial. *Computer*, 29(3):31–44.

[Janouchová et al., 2021] Janouchová, J., Kučerová, A., Sýkora, J., Vorel, J., and Wan-Wendner, R. (2021). Robust probabilistic calibration of a stochastic lattice discrete particle model for concrete. *Engineering Structures*, 236:112000.

[Janssen and Heuberger, 1995] Janssen, P. and Heuberger, P. (1995). Calibration of process-oriented models. *Ecological Modelling*, 83(1):55–66.

[Jawaid et al., 2018] Jawaid, M., Thariq, M., and Saba, N. (2018). *Structural Health Monitoring of Biocomposites, Fibre-Reinforced Composites and Hybrid Composites*. Elsevier Science.

[Jaynes, 1957] Jaynes, E. (1957). Information theory and statistical mechanics. *Physical review*, 106(4):620.

[Jeffreys, 1946] Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453–461.

[Jennrich and Schluchter, 1986] Jennrich, R. and Schluchter, M. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, 42(4):805–820.

[Jiang et al., 2010] Jiang, X., Mahadevan, S., and Urbina, A. (2010). Bayesian nonlinear structural equation modeling for hierarchical validation of dynamical systems. *Mechanical Systems and Signal Processing*, 24(4):957–975.

[Johnson, 2007] Johnson, S. G. (2007). The NLopt nonlinear-optimization package. https://github.com/stevengj/nlopt.

[Jolling et al., 2005] Jolling, K., Ruixo, J., Hemeryck, A., Vermeulen, A., and Greway, T. (2005). Mixed-effects modelling of the interspecies pharmacokinetic scaling of pegylated human erythropoietin. *European journal of pharmaceutical sciences*, 24(5):465–475.

[Julien, 2010] Julien, C. (2010). *Conception Optimale de l'Anisotropie dans les Structures Stratifiées à Rigidité Variable par la Méthode Polaire-Génétique (in french)*. Ph.d. thesis, ED-488 SIS.

[Jung et al., 2022] Jung, Y., Jo, J., Choo, J., and Lee, I. (2022). Statistical model calibration and design optimization under aleatory and epistemic uncertainty. *Reliability Engineering & System Safety*, 222:108428.

[Kalman, 1960] Kalman, R. (1960). A new approach to linear filtering and prediction problems. *ASME Journal of Basic Engineering*, 82:35–45.

[Kariya and Kurata, 2004] Kariya, T. and Kurata, H. (2004). *Generalized Least Squares*. John Wiley & Sons.

[Kennedy and O'Hagan, 2001] Kennedy, M. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464.

[Kimand and Youn, 2019] Kimand, T. and Youn, B. (2019). Identifiability-based model decomposition for hierarchical calibration. *Structural and Multidisciplinary Optimization*, 60:1801–1811.

[Kristensen et al., 2016] Kristensen, K., Nielsen, A., Berg, C., Skaug, H., and Bell, B. (2016). Tmb: Automatic differentiation and laplace approximation. *Journal of Statistical Software*, 70(5).

[Kruschke, 2014] Kruschke, J. (2014). Doing bayesian data analysis: A tutorial with r, jags, and stan.

[Kuhn and Lavielle, 2004] Kuhn, E. and Lavielle, M. (2004). Coupling a stochastic approximation version of em with a mcmc procedure. *Probability and Statistics*, 8:115–131.

[Kullback and Leibler, 1951] Kullback, S. and Leibler, R. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79 – 86.

[Kunstner et al., 2019] Kunstner, F., Hennig, P., and Balles, L. (2019). Limitations of the empirical fisher approximation for natural gradient descent. *Advances in neural information processing systems*, 32.

[Laboulfie et al., 2021] Laboulfie, C., Balesdent, M., Brevault, L., Irisarri, F.-X., Maire, J.-F., Da Veiga, S., and Le Riche, R. (2021). Calibration of material model parameters using mixed-effects models. In *4th International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2021)*, pages 258–295, Athens, Greece.

[Laboulfie et al., 2022] Laboulfie, C., Balesdent, M., Brevault, L., Irisarri, F.-X., Maire, J.-F., Da Veiga, S., and Le Riche, R. (2022). Sequential calibration of material model using mixed-effects. In *25ème Congrès Français de Mécanique*, Nantes, France.

[Laboulfie et al., 2023] Laboulfie, C., Balesdent, M., Brevault, L., Irisarri, F.-X., Maire, J.-F., Da Veiga, S., and Le Riche, R. (2023). Sequential calibration of material constitutive model using mixed-effects. *Mechanics & Industry*, 24:27.

[Laird and Ware, 1982] Laird, N. and Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974.

[Laurin, 2005] Laurin, F. (2005). *Approche Multimodèle des Mécanismes de Ruine Progressive des Matériaux Stratifiés et Analyse de la Tenue de Structures Composites (in french).* Ph.d. thesis, Univerité de Franche-Comté, Onera.

[Lavielle, 2008] Lavielle, M. (2008). Monolix (modèles non linéaires à effets mixtes) user guide.

[Lavielle, 2014] Lavielle, M. (2014). *Mixed-Effects Models for the Population Approach: Models, Tasks, Methods and Tools.* Chapman.

[Lee, 2022] Lee, S. (2022). Bayesian nonlinear models for repeated measurement data: An overview, implementation, and applications. *Mathematics*, 10(6):898.

[Lekou and Philippidis, 2008] Lekou, D. and Philippidis, T. (2008). Mechanical property variability in frp laminates and its effect on failure prediction. *Composites Part B: Engineering*, 39(7-8):1247–1256.

[Levenberg, 1944] Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164–168.

[Lewis and Beal, 1918] Lewis, J. and Beal, C. H. (1918). Some new methods for estimating the future production of oil wells. *Transactions of the AIME*, 59(1):492–525.

[Lindstrom and Bates, 1990] Lindstrom, M. and Bates, D. (1990). Nonlinear mixed-effects models for repeated measures data. *Biometrics*, 46(3):673–687.

[Link and Eaton, 2012] Link, W. and Eaton, M. (2012). On thinning of chains in mcmc. *Methods in Ecology and Evolution*, 3(1):112–115.

[Liu, 1988] Liu, R. (1988). Bootstrap procedures under some non-i.i.d. models. *The Annals of Statistics*, 16(4):1696– 1708.

[Loeve, 1978] Loeve, M. (1978). *Probability Theory II.* Springer Science & Business Media.

[Longford, 1987] Longford, N. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 74(4):817–827.

[Louis, 1982] Louis, T. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(2):226–233.

[Lu et al., 1997] Lu, J.-C., Park, J., and Yang, Q. (1997). Statistical inference of a time-to-failure distribution derived from linear degradation data. *Technometrics*, 39(4):391–400.

[Ly et al., 2007] Ly, A., Marsman, M., Verhagen, J., Grasman, R., and Wagenmakers, E. (2007). A tutorial on fisher information. *Journal of Mathematical Psychology*, 80:40–55.

[Madsen, 2003] Madsen, H. (2003). Parameter estimation in distributed hydrological catchment modeling using automatic calibration with multiple objectives. *Advances in Water Resources*, 26(2):205–216.

[Maire et al., 1997] Maire, J.-F., Lesne, M., and Girard, R. (1997). An explicit behavioural damage model for the design of components in ceramic matrix composites. *Key Engineering Materials*, 127-131:1053–1060.

[Maljaars et al., 2022] Maljaars, J., Rózsás, A., Walters, C., and Slot, H. (2022). Uncertainty quantification of the failure assessment diagram for flawed steel components in bs 7910:2019. *Engineering Fracture Mechanics*, 268:108446.

[Mallet et al., 1998] Mallet, A., Mentré, F., Steimer, J.-L., and Lokiec, F. (1998). Nonparametric maximum likelihood estimation for population pharmacokinetics, with application to cyclosporine. *Journal of Pharmacokinetics and Biopharmaceutics*, 16(3):311–327.

[Mammen, 1992] Mammen, E. (1992). *When Does Bootstrap Work?: Asymptotic Results and Simulations*. Lecture notes in statistics. Springer-Verlag.

[Marchand et al., 2016] Marchand, B., Chamoin, L., and Rey, C. (2016). Real-time updating of structural mechanics models using kalman filtering, modified constitutive relation error, and proper generalized decomposition. *International Journal for Numerical Methods in Engineering*, 107(9):786–810.

[Marcin et al., 2011] Marcin, L., Maire, J.-F., Carrère, N., and Martin, E. (2011). Development of a macroscopic damage model for woven ceramic matrix composites. *International Journal of Damage Mechanics*, 20(6):939–957.

[Marquardt, 1963] Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameter. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441.

[Marquez Costa, 2021] Marquez Costa, J.-P. (2021). *Caractérisation et modélisation des interfaces dans les composites organiques stratifiés à haute température : Application à la tenue au feu des structures aéronautiques (in french)*. Ph.d. thesis, ISAE-ENSMA Ecole Nationale Supérieure de Mécanique et d'Aérotechique - Poitiers.

[Martens, 2020] Martens, J. (2020). New insights and perspectives on the natural gradient method. *Journal of Machine Learning Research*, 21(1):1–76.

[Marti et al., 2010] Marti, R., Moreno-Vega, J., and Duarte, A. (2010). *Advanced Multi-start Methods*, pages 265–281.

[Martin and Simpson, 2005] Martin, J. and Simpson, T. (2005). Use of kriging models to approximate deterministic computer models. *AIAA Journal*, 43(4):853–863.

[McFarland et al., 2008] McFarland, J., Mahadevan, S., and Romeroand, V. Swiler, L. (2008). Calibration and uncertainty analysis for computer simulations with multivariate output. *AIAA Journal*, 46(5):1253–1265.

[Mentré et al., 1997] Mentré, F., Mallet, A., and Baccar, D. (1997). Optimal design in random-effects regression models. *Biometrika*, 84(2):429–442.

[Meraghni et al., 2014] Meraghni, F., Chemisky, Y., Piotrowski, B., Echchorfi, R., Bourgeois, N., and Patoor, E. (2014). Parameter identification of a thermodynamic model for superelastic shape memory alloys using analytical calculation of the sensitivity matrix. *European Journal of Mechanics-A/Solids*, 45:226–237.

[Metropolis et al., 1953] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

[Metropolis and Ulam, 1949] Metropolis, N. and Ulam, S. (1949). The monte carlo method. *Journal of the American statistical association*, 44(247):335–341.

[Molimard and Le Riche, 2003] Molimard, J. and Le Riche, R. (2003). Identification de piézoviscosité en lubrification (in french). *Mechanics & Industry*, 4(6):645–653.

[Molimard et al., 2005] Molimard, J., Le Riche, R., Vautrin, A., and Lee, J. (2005). Identification of the Four Orthotropic Plate Stiffnesses Using a Single Open-hole Tensile Test (in french). *Experimental Mechanics*, 45(5):404–411.

[Morris, 1991] Morris, M. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174.

[Murugan et al., 2008] Murugan, S., Harursampath, D., and Ganguli, R. (2008). Material uncertainty propagation in helicopter nonlinear aeroelastic response and vibratory analysis. *AIAA Journal*, 46(9):2332–2344.

[Nagel and Sudret, 2016] Nagel, J. and Sudret, B. (2016). A unified framework for multilevel uncertainty quantification in Bayesian inverse problems. *Probabilistic Engineering Mechanics*, 43:68–84.

[Naslain, 2004] Naslain, R. (2004). Design, preparation and properties of non-oxide cmcs for application in engines and nuclear reactor: an overview. *Composites Science and Technology*, 64(2):155–170.

[Nelson and Siegel, 1987] Nelson, C. and Siegel, A. (1987). Parsimonious modeling of yield curves. *The Journal of Business*, 60:473–489.

[Nelson, 2004] Nelson, W. (2004). *Accelerated testing: statistical models, test plans and data analysis*. John Wiley & Sons.

[Nguyen and Mentré, 2014] Nguyen, T. and Mentré, F. (2014). Evaluation of the fisher information matrix in nonlinear mixed-effect models using adaptive gaussian quadrature. *Computational Statistics & Data Analysis*, 80:57–69.

[Niyigena et al., 2016] Niyigena, C., Amziane, S., Chateauneuf, A., Arnaud, L., Bessette, L., Collet, F., Lanos, C., Escadeillas, G., Lawrence, M., Magniont, C., Marceau, S., Pavia, S., Peter, U., Picandet, V., Sonebi, M., and Walker, P. (2016). Variability of the mechanical properties of hemp concrete. *Materials Today Communications*, 7:122–133.

[Nocedal and Wright, 2006] Nocedal, J. and Wright, S. (2006). *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer New York.

[Nuismer and Whitney, 1975] Nuismer, R. and Whitney, J. (1975). Uniaxial failure of composite laminates containing stress concentrations. *ASTM special technical publications*, pages 117–142.

[Oberkampf and Roy, 2010] Oberkampf, W. and Roy, C. (2010). *Verification and Validation in Scientific Computing*. Cambridge University Press.

[Ogden, 2021] Ogden, H. (2021). On the error in laplace approximations of high-dimensional integrals. *Stat*, 10(1):e380.

[Pacheco et al., 2016] Pacheco, C., Dulikravich, G., Vesenjak, M., Borovinšek, M., Duarte, I., Jha, R., Reddy, S., O., H., and Colaco, M. (2016). Inverse parameter identification in solid mechanics using bayesian statistics, response surfaces and minimization. *Technische Mechanik*, 36(1-2):120–131.

[Papadimas and Dodwell, 2021] Papadimas, N. and Dodwell, T. (2021). A hierarchical bayesian approach for calibration of stochastic material models. *Data-Centric Engineering*, 2(0):e20.

[Paranjape et al., 2021] Paranjape, H., Aycock, K., Bonsignore, C., Weaver, J., Craven, B., and Duerig, T. (2021). A probabilistic approach with built-in uncertainty quantification for the calibration of a superelastic constitutive model from full-field strain data. *Computational Materials Science*, 192(0):110357.

[Pardo, 2018] Pardo, L. (2018). *Statistical inference based on divergence measures*. Chapman and Hall/CRC.

[Pinheiro, 2022] Pinheiro, J. (2022). Linear and nonlinear mixed-effects models. https://cran.r-project.org/web/packages/nlme/nlme.pdf.

[Pinheiro and Bates, 1994] Pinheiro, J. and Bates, D. (1994). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4(1):12–35.

[Pinheiro and Bates, 1996] Pinheiro, J. and Bates, D. (1996). Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, 6:289–296.

[Pinheiro and Bates, 2002] Pinheiro, J. and Bates, D. (2002). *Mixed-Effect Models in S and S-plus*, volume 96.

[Pinheiro et al., 2001] Pinheiro, J., Liu, C., and Wu, Y. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics*, 10(2):249–276.

[Pintelon and Schoukens, 2012] Pintelon, R. and Schoukens, J. (2012). *System Identification*. John Wiley & Sons.

[Plan et al., 2012] Plan, E., Maloney, A., Mentré, F., Karlsson, M., and Bertrand, J. (2012). Performance comparison of various maximum likelihood nonlinear mixed-effects estimation methods for dose-response models. *The AAPS journal*, 14:420–432.

[Pollak and Palazotto, 2009] Pollak, R. and Palazotto, A. (2009). A comparison of maximum likelihood models for fatigue strength characterization in materials exhibiting a fatigue limit. *Probabilistic Engineering Mechanics*, 24(2):236–241.

[Qiu et al., 2002] Qiu, Z., Song, P., and Tan, M. (2002). Bayesian hierarchical models for multi-level repeated ordinal data using winbugs. *Journal of Biopharmaceutical Statistics*, 12(2):121–135.

[Rajib et al., 2016] Rajib, M., Merwade, V., and Yu, Z. (2016). Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture. *Journal of Hydrology*, 536:192–207.

[Rana and Fangueiro, 2016] Rana, S. and Fangueiro, R. (2016). Advanced composites in aerospace engineering. In *Advanced composite materials for aerospace engineering*, pages 1–15. Elsevier.

[Rappel et al., 2019a] Rappel, H., Beex, L., and Bordas, S. (2019a). Bayesian inference to identify parameters in viscoelasticity. *Mechanics of Time-Dependent Materials*, 22:221–258.

[Rappel et al., 2019b] Rappel, H., Beex, L., Hale, J., Noels, L., and Bordas, S. (2019b). A tutorial on bayesian inference to identify material parameters in solid mechanics. *Archives of Computational Methods in Engineering*, 27:361–385.

[Riviere et al., 2016] Riviere, M.-K., Ueckert, S., and Mentré, F. (2016). An mcmc method for the evaluation of the fisher information matrix for non-linear mixed-effect models. *Biostatistics*, 17(4):737–750.

[Robert and Casella, 2005] Robert, C. and Casella, G. (2005). *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag.

[Rollet, 2007] Rollet, Y. (2007). *Vers une Maîtrise des Incertitudes en Calcul des Structures Composites (in french)*. Ph.d. thesis, É'cole Polytechnique.

[Rosbjerg and Madsen, 2006] Rosbjerg, D. and Madsen, H. (2006). *Concepts of Hydrologic Modeling*, chapter 10. John Wiley & Sons.

[Rosenblatt, 1952] Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472.

[Rosenkrantz, 1989] Rosenkrantz, R. (1989). *Prior Probabilities*, pages 114–130. Springer Netherlands.

[Rouder and Lu, 2005] Rouder, J. and Lu, J. (2005). An introduction to bayesian hierarchical models with an application in theory of signal detection. *Psychonomic bulletin & review*, 12(4):573–604.

[Sankararaman et al., 1991] Sankararaman, S., Ling, Y., and Mahadevan, S. (1991). Uncertainty quantification and model validation of fatigue crack growth prediction. *Engineering Fracture Mechanics*, 78(7):1487–1504.

[Savic and Karlsson, 2009] Savic, R. and Karlsson, M. (2009). Importance of shrinkage in empirical bayes estimates for diagnostics: Problems and solutions. *The AAPS journal*, 11:558–69.

[Schepsmeier and Stöber, 2014] Schepsmeier, U. and Stöber, J. (2014). Derivatives and fisher information of bivariate copulas. *Statistical Papers*, 55:35.

[Schuurman et al., 2016] Schuurman, N., Grasman, R., and Hamaker, E. (2016). A comparison of inverse-wishart prior specifications for covariance matrices in multilevel autoregressive models. *Multivariate Behavioral Research*, 51(2-3):185–206.

[Schwarz, 1978] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

[Schwer, 2006] Schwer, L. (2006). *Guide for Verification and Validation in Computational Solid Mechanics*. ASME.

[Shannon, 1948] Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(1):379–423.

[Sheiner and Beal, 1981] Sheiner, L. and Beal, S. (1981). Evaluation of methods for estimating population pharmacokinetic parameters ii. biexponential model and experimental pharmacokinetic data. *Journal of pharmacokinetics and biopharmaceutics*, 9(5):635–651.

[Sinharay, 2003] Sinharay, S. (2003). Assessing convergence of the markov convergence of the markov chain monte carlo algorithms: a review. *ETS Research Report Series*, 2003(1):1–52.

[Sklar, 1959] Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges (in french). In *Annales de l'ISUP*, volume 8, pages 229–231.

[Smarslok et al., 2012] Smarslok, B., Haftka, R., and Ifju, P. (2012). Correlation model for composite elastic properties using only measurements from a single specimen. *Probabilistic Engineering Mechanics*, 29(0):64–69.

[Smolyak, 1963] Smolyak, S. (1963). Quadrature and interpolation formulas for tensor products of certain classes of functions. In *Doklady Akademii Nauk*, volume 148, pages 1042–1045. Russian Academy of Sciences.

[Sobol, 2001] Sobol, I. (2001). Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280.

[Song et al., 2020] Song, M., Behmanesh, I., Moaveni, B., and Papadimitriou, C. (2020). Accounting for modeling errors and inherent structural variability through a hierarchical bayesian model updating approach: An overview. *Sensors*, 20(14):3874–3901.

[Song et al., 2005] Song, P.-K., Fan, Y., and Kalbfleisch, J. (2005). Maximization by parts in likelihood inference. *Journal of the American Statistical Association*, 100(472):1145–1158.

[Sriramula and Chryssanthopoulos, 2009] Sriramula, S. and Chryssanthopoulos, M. (2009). Quantification of uncertainty modelling in stochastic analysis of frp composites. *Composites Part A: Applied Science and Manufacturing*, 40(11):1673–1684.

[Stegmann et al., 2007] Stegmann, G., Jacobucci, R., Harring, J., and Grimm, K. (2007). Nonlinear mixed-effects modeling programs in R. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(1):160–165.

[Stram and Lee, 1994] Stram, D. and Lee, J. (1994). Variance components testing in the longitudinal mixed-effects model. *Biometrics*, pages 1171–1177.

[Svensson, 1994] Svensson, L. (1994). Estimating and Interpreting Forward Interest Rates: Sweden 1992 - 1994. Working Papers 4871, National Bureau of Economic Research.

[Swanson et al., 1985] Swanson, S., Messick, M., and Toombes, G. (1985). Comparison of torsion tube and iosipescu in-plane shear test results for a carbon fibre-reinforced epoxy composite. *Composites*, 16(3):220–224.

[Tan, 2017] Tan, S. (2017). *Stress Concentrations in Laminated Composites*. Routledge.

[Tanner, 1998] Tanner, M. (1998). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions, 3rd Edition*, volume 54 of *Lecture Notes in Statistics*. Springer New York.

[Tarantola, 2005] Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics.

[Thai et al., 2013] Thai, H., Mentré, F., Holford, N., Veyrat-Follet, C., and Comets, E. (2013). A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models. *Pharmaceutical statistics*, 12.

[Thai et al., 2014] Thai, H.-T., Mentré, F., Holford, N., Veyrat-Follet, C., and Comets, E. (2014). Evaluation of bootstrap methods for estimating uncertainty of parameters in nonlinear mixed-effects models: a simulation study in population pharmacokinetics. *Journal of Pharmacokinetics and Pharmacodynamics*, 41(1):15 – 33.

[Tibshirani, 2016] Tibshirani, R. (2016). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

[Tikhonov et al., 1995] Tikhonov, A., Goncharsky, A., Stepanov, V., and Yagola, A. (1995). *Numerical Methods for the Solution of Ill-Posed Problems*. Mathematics and Its Applications. Springer Netherlands.

[Toft et al., 2013] Toft, H., Branner, K., Mishnaevsky, L. J., and Sørensen, J. (2013). Uncertainty modelling and code calibration for composite materials. *Journal of Composite Materials*, 47(14):1729–1747.

[Tomblin et al., 2001] Tomblin, J., Ng, Y., and Keshavanarayana, S. (2001). Material qualification and equivalency for polymer matrix composite material systems. Technical report.

[Urbina et al., 2012] Urbina, A., Paez, T., and Consulting, T. (2012). A bayes network approach to uncertainty quantification in hierarchically-developed computational models. *International Journal for Uncertainty Quantification*, 2(2).

[Vaida and Blanchard, 2005] Vaida, F. and Blanchard, S. (2005). Conditional akaike information for mixed-effects models. *Biometrika*, 92(2):351–370.

[Van Veldhuizen, 1999] Van Veldhuizen, D. (1999). *Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations*. Storming Media.

[Verbeke and Molenberghs, 2013] Verbeke, G. and Molenberghs, G. (2013). The gradient function as an exploratory goodness-of-fit assessment of the random-effects distribution in mixed models. *Biostatistics (Oxford, England)*, 14(3):477–490.

[Vinot et al., 2023] Vinot, M., Liebold, C., Usta, T., Holzapfel, M., Toso, N., and Voggenreiter, H. (2023). Stochastic modeling of continuous glass-fiber reinforced plastics–considering material uncertainty in microscale simulations. *Journal of Composite Materials*, 57(1):133–145.

[Viricelle et al., 2001] Viricelle, J., P., G., and D., B.-H. (2001). Oxidation behaviour of a multi-layered ceramic-matrix composite (sic)f/c/(sibc)m. *Composites Science and Technology*, 61(4):607–614.

[Vonesh et al., 2002] Vonesh, E., Wang, H., Nie, L., and Majumdar, D. (2002). Conditional second-order generalized estimating equations for generalized linear and nonlinear mixed-effects models. *Journal of the American Statistical Association*, 97(457):271–283.

[Vong et al., 2014] Vong, C., Ueckert, S., Nyberg, J., and Hooker, A. (2014). Handling below limit of quantification data in optimal trial design. *Journal of Pharmacokinetics and Pharmacodynamics*.

[Vrugt et al., 2003] Vrugt, J., Gupta, H., Bastidas, L. A., Bouten, W., and Sorooshian, S. (2003). Effective and efficient algorithm for multiobjective optimization of hydrologic models: multiobjective optimization of hydrologic models. *Water Resources Research*, 39(8).

[Wakefield, 1996] Wakefield, J. (1996). The bayesian analysis of population pharmacokinetic models. *Journal of the American Statistical Association*, 91(433):62–75.

[Walter et al., 1997] Walter, E., Pronzato, L., and Norton, J. (1997). *Identification of parametric models from experimental data*, volume 1. Springer.

[Wand and Jones, 1994] Wand, M. and Jones, M. (1994). *Kernel Smoothing*. Taylor & Francis.

[Wang et al., 2020] Wang, P., Zheng, X., Li, J., and Zhu, B. (2020). Prediction of epidemic trends in covid-19 with logistic model and machine learning technics. *Chaos, Solitons & Fractals*, 139:110058.

[Wasserman, 2010] Wasserman, L. (2010). *All of Statistics: A Concise Course in Statistical Inference*. Springer Publishing Company, Incorporated.

[Wei and Tanner, 1990] Wei, G. and Tanner, M. (1990). A monte carlo implementation of the em algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, 85(411):699–704.

[Wolfinger, 1993] Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika*, 80(4):791–795.

[Wong et al., 2017] Wong, R., Storlie, C., and Lee, T. (2017). A frequentist approach to computer model calibration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79:635–648.

[Wu, 1983] Wu, C. (1983). On the convergence properties of the em algorithm. *The Annals of Statistics*, 11(1):95–103.

[Wu and Zhang, 2002] Wu, H. and Zhang, J.-T. (2002). The study of long-term hiv dynamics using semi-parametric non-linear mixed-effects models. *Statistics in Medicine*, 21(23):3655–3675.

[Wu et al., 2018] Wu, S., Angelikopoulos, P., Beck, J., and Koumoutsakos, P. (2018). Hierarchical stochastic model in bayesian inference for engineering applications: Theoretical implications and efficient approximation. *ASCE-ASME J. Risk and Uncertainty in Engineering Systems, Part B: Mech Engrg*, 5(1).

[Ye et al., 2022] Ye, J., Mahmoudi, M., Karayagiz, K., Johnson, L., Seede, R., Karaman, Y., Arroyave, R., and Elwany, A. (2022). Bayesian calibration of multiple coupled simulation models for metal additive manufacturing: A bayesian network approach. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, 8(1):12.

[Youn et al., 2011] Youn, B., Jung, B., Xi, Z., Kim, S., and Lee, W. (2011). A hierarchical framework for statistical model calibration in engineering product development. *Computer Methods in Applied Mechanics and Engineering*, 200(13):1421–1431.

[Zawada et al., 2017] Zawada, L. P., Pierce, J. L., Przybyla, C. P., and Katz, A. P. (2017). Evaluation of a melt infiltrated sic/sic ceramic matrix composite. Technical report, Air Force Research Laboratory Wright Patterson Air Force Base United States.

ABSTRACT

Many model calibration techniques take into account the impact of measurement errors, but more rarely that of material variability on model parameters, which is significant in the case of composite materials. The aim of this work is to develop a calibration method to estimate the influence of variability on the parameters of complex composite behavior laws. The statistical framework of usual methods for quantifying uncertainties is first analyzed in a literature review. The application of these methods to repetitions of monotonic tensile tests illustrates the importance of material variability and the limitations of conventional methods for estimating it. To characterize its impact on material properties, the framework of population approaches, already implemented in structural mechanics problems, is adopted. We focus in particular on mixed-effects models, characterized by their considerable flexibility of use. These methods are used to demonstrate their relevance on an elementary model, and test their settings and assumptions. These methods are used to demonstrate their relevance to an elementary model, and test their settings and hypotheses. Two points are specifically studied: the number of individuals, a key feature to guarantee confidence in material variability estimates, and the complexity of variability modeling (in particular the interest of taking into account statistical dependencies between model parameters). This study is carried out on a simple uniaxial damageable elastic behavior model with four parameters to be identified. To tackle more advanced mechanical models, we propose a sequential calibration strategy, compatible with mixed-effects models. The method, which is applied to synthetic and then real data, allows, compared with conventional mixed-effects, to estimate material variability with equivalent accuracy, while significantly reducing the number of model runs. Finally, the calibrated distributions are propagated to estimate the failure stress of a perforated composite plate in tension, thus illustrating the entire uncertainty quantification chain, from estimation to propagation.

Keywords : calibration of composites behavior laws, variability, mixed-effects models.

RÉSUMÉ (IN FRENCH)

De nombreuses techniques de calibration de modèle permettent de prendre en compte l'impact des erreurs de mesure, mais plus rarement celui de la variabilité matériau sur les paramètres de modèle, celle-ci étant significative dans le cas des composites. L'objectif de ce travail est de mettre en place une méthode de calibration pour estimer l'influence de la variabilité sur les paramètres de lois complexes de comportement composite. Le cadre d'application statistique des méthodes usuelles de quantification d'incertitudes est d'abord analysé dans une revue bibliographique. L'application de ces méthodes à des répétitions d'essais de traction monotone illustre l'importance de la variabilité matériau et les limites des méthodes usuelles pour son estimation. Pour caractériser son impact sur les propriétés du matériau, le cadre des approches par population, déjà mis en œuvre dans des problématiques de mécanique des structures, est adopté. On se concentre en particulier sur les modèles à effets-mixtes, caractérisés par une importante souplesse d'utilisation. Ces méthodes sont mises en œuvre pour démontrer leur pertinence sur un modèle élémentaire, et tester leurs réglages et hypothèses. Deux points sont spécifiquement étudiés : le nombre d'individus, déterminant pour garantir la confiance sur les estimations de la variabilité matériau, et la complexité de la modélisation de la variabilité (notamment la prise en compte des dépendances statistiques entre paramètres de modèle). Cette étude est réalisée sur un modèle de comportement élastique endommageable uniaxial simple avec quatre paramètres à identifier. Pour aborder des modèles mécaniques plus avancés, nous proposons une stratégie de calibration séquentielle, compatible avec les modèles à effets-mixtes. La méthode, appliquée sur des données synthétiques puis réelles, permet, par rapport à une application classique des effets-mixtes, d'estimer avec une précision équivalente la variabilité matériau en divisant environ par deux le nombre d'appels au modèle. Enfin, les distributions calibrées sont propagées pour estimer l'effort à rupture d'une plaque composite perforée, sollicitée en traction, ce qui illustre l'ensemble de la chaîne de quantification d'incertitudes, depuis leur estimation jusqu'à leur propagation.

Mots-clés : calibration de lois de comportement composites, variabilité, modèles à effets-mixtes.