Aleksander Sinek

Integration of Statistical Data Analysis and Surrogate Modeling for Uncertainty Quantification, Sensitivity Analysis and Inverse Problems involving Fluid-Structure Interaction models

Doctoral dissertation

Supervisors:
Prof. Wojciech Adamczyk

Prof. Leif Rune Hellevik

Science discipline:
Environmental Engineering, Mining and Energy

This research is supported by National Science Centre Poland within projects UMO-2018/31/B/ST8/02201. This help is gratefully acknowledged herewith.

Silesian University of Technology

Gliwice, 2024

Author
Aleksander Sinek
Silesian University of Technology
Faculty of Energy and Environmeental Engineering
Department of Thermal Technology
Konarskiego St. 22
44-100 Gliwice
Poland

Norwegian University of Science and Technonlogy
Faculty of Engineering
Department of Structural Engineering
Richard Birkelands vei 1A
7034 Trondheim
Norway

e-mail (academic): aleksander.sinek@polsl.pl ; alekstsi@stud.ntnu.no

Acknowledgments

First of all I would like to thank my supervisors Wojciech Adamczyk and Leif Rune Hellvik for guiding me throughout my thesis, having endless patience and always supporting me. My thanks also go out to Jacob Sturdy for his continued support and many insightful conversations about statistics that inspired many of the topics covered. My thanks also go out to Mateusz Mesek, Marek Rojczyk and Jan Juszczyk for always having my back and working with me tirelessly in the lab. I would also like to thank Ryszard Białecki and Ziemowit Ostrowski for helping me navigate the administrative hellscape of a PhD and providing professional feedback on scientific writing, analyses and more. I would also like to thank Wojciech Wolański and Grzegorz Kokot for performing material tests of arterial phantoms used in this work.

My thanks also go out to all the fellow (ex) PhD students of the Department of Thermal Technology (Katedra Techniki Cieplnej) at the Silesian University of Technology (Politechnika Śląska) in Gliwice who had a more or less direct impact on my thesis: Marcin Nowak, Agata Widuch, Edyta Piechnik, Jakub Chrobak. Thanks also go out to other employees and scientists: Jacek Smołka, Arkadiusz Ryfa, Adam Klimanek and others.

Similarly I would like to thank all my fellow (ex) PhD Students at the Department of Structural Engineering at the Norwegian Univesrity of Science and Technology in Trondheim: Friederike Elisabeth Schafer, Erling Velten Rothmund, Nikolai Bjordalsbakke, Frederik Eikeland Fossan, Radha Krishna Ayyalasomayajula, Javad Sadeghinia, Omid Ghorbani and Stefan Gotz. My thanks also go out to the employees and scientists at the depratment: Victorien Prot, Bjorn Skallerud, Anders Lervik and others.

Personally my thanks go out to my family, particularly my parents Tatiana Sinek and Sebastian Sinek for always being there and to my younger brother, Karol Sinek for always asking insightful questions. Moreover, my thanks go to

closest friends and other people who have had an influence on me during this period of time. If you were not mentioned by name, such is the nature of short acknowledgments, no-hard feelings and thanks nonetheless!

Last, but not least I would like to thank my fiancee, Agnieszka Ciesielska, for the never-ending support and patience during my entire PhD adventure.

This scientific work in the presented form would not have been possible without the financial assistance of the National Science Center (Poland) project "ENTHRAL - Non-invasive in-vivo assessment of stiffness of human artery walls" under grant$\#UMO-2019/34/H/ST8/00624$ funded by the Norwegian Financial Mechanism 2014-2021. The support is gratefully acknowledged herewith.

Contents

List of Figures

List of Tables

Abbreviation	Meaning
ANN	Artifical Neural Networks
ANOVA	Analysis of Variance
CFD	Computational Fluid Dynamics
CI	Confidence Interval
DOE	Design Of Experiments
FEA	Finite Element Analysis
FOM	Full Order Model
FSI	Fluid-Structure Interaction
GP	Gaussian Process
GPR	Gaussian Process Regression
HDMR	High Dimensional Model Representation
HMC	Hamiltonian Monte Carlo
IP	Inverse problem
KR	Kernel Regression
LHS	Latin Hypercube Sampling
MAE	Mean Absolute Error
MaxErr	Maximum Absolute Error
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
MEDS	Median Absolute Error
MLE	Maximum Likelihood Estimation
PCA	Principal Component Analysis
PCE	Polynomial Chaos Expansion
PI	Prediction Interval
PPC	Posterior/Prior predictive check
PPD	Posterior Predictive Distribution
RMSE	Root Mean-Squared Error
ROM	Reduced Order Model
RV	Random Variable
SA	Sensitivity Analysis
SGPR	Sparse Gaussian Process Regression
SVGPR	Sparse Variation Gaussian Process Regression
SM	Surrogate Model
SVR	Support Vector Regression
UQ	Uncertainty Quantification

1 Introduction

1.1 Medical Background and contemporary methodologies

According to World Health Federation (WHF) 2023 Heart Report [1], Cardiovacular Diseases (CVDs) are still the leading cause of death among humans. Cardiovascular Diseases are any diseases that are responsible for medical pathologies related to the heart or blood vessels. CVDs are a class of diseases and medical conditions themselves, which constitute f.ex. coronary artery diseases, hypertensive heart diseases, ischemic cerebrovascular accident (strokes) and many others [2]. In 2021, their death toll was estimated at 20.5 million, which constitutes a third of all deaths globally. Despite the growing health awareness and slowly decreasing rates of CVD-related deaths in the recent years, they still pose a substantial threat. Even though there are a lof of treatments possible, unfortunately, there is no universal approach for (early) detection and, sadly, it most often occurs after an cardiac episode has already taken place. One of the steps toward raising a CVD-related concern in a patient is the determination of their overall risk factor associated with CVDs, which consists of many smaller factors contributing negatively to overall well-being:

elevated LDL cholesterol, sodium intake, air pollution, high body-mass index, tobacco use, low physical activity, elevated blood pressure (hypertension) and others. Each of these factors can be determined through a thorough interview conducted by a medical examiner. Although most of these factors are readily available to physicians, they are not sufficient to constitute a basis for taking a serious medical or surgical action. Furthermore, they do not explain the underlying mechanism behind many of the reported CVDs, which is arterial stiffness. [3] Arterial stiffening is a process of decreasing arterial compliance as a result of various factors, primarily aging and arteriosclerosis [2]. This loss of compliance has many negative outcomes. First, arteries serve as a secondary pumping mechanism, i.e. when the heart ejects blood, the arteries expand and gain elastic potential energy. After a second they bounce back to their original shape, transforming the elastic energy into work, and pumping the accumulated blood down along the arterial tree. An impediment to this mechanism leads, for example, to left ventricular hypertrophy, left ventricular remodeling, and ultimately to heart failure [4]. An increased workload on the heart often leads to hypertension and strokes [4]. As demonstrated, arterial stiffness correlates with many CVDs and thus it would be beneficial to have a way for physicians to accurately estimate it so that appropriate treatment or lifestyle changes may be recommended. The current gold standard of arterial stiffness measurement is the Pulse Wave Velocity measurement. The mathematical formula itself has many variants depending on the application, however the most widely known version is the Moens-Korteweg relation: [3]

$$PWV = \sqrt{\frac{E \cdot h}{2 \cdot b \cdot \rho}} \tag{1.1}$$

where ρ is the blood density, b is the blood vessel radius, h is the vessel wall thickness and E is the Young's modulus, defined as: [3]

$$E = \frac{\sigma(\epsilon)}{\epsilon} \tag{1.2}$$

where σ is the stress tensor and ϵ is the strain tensor. The Pulse Wave Velocity is obtained by measuring the Pulse Wave at two locations in the body, most often carotid and femoral arteries, and the equation is reordered to obtain the arterial stiffness, E. This is a so called 'regional' or 'global' method, as it estimates arterial stiffness using data and measurements taken at different, distant

regions within a human body, which increases the inaccuracy of estimation. PWV-based methods generally are widespread and inexpensive, however they are best suited for recovering the stiffness in larger vessels. [5] Furthermore, PWV-based methods suffer from impaired diagnostic capabilities resulting from the fact that they do not take into account local inhomogeneities in stiffness within the arterial tree. This omission is crucial, as in the beginning stages of various atherosclerotic and arteriosclerotic diseases, only the local elastic properties of the arterial walls are affected. Thus they may be easily missed by such a regional approach. [6] Hence there exists a need for a framework that will provide accurate, fast and local estimates of arterial stiffness for early detection of onset of arterial stiffening, so that CVD-preventative measures may be advised before any real pathologies develop. This thesis proposes an alternative methodology relying on the techniques introduced in the following sections. Once the methodology and objectives are introduced, the techniques are briefly described.

1.2 Obejctives and Methodology

As mentioned in the previous section, there currently exists a gap in the literature, when it comes to an efficient method for reliably estimating local arterial stiffness. The purpose of the ENTHRAL project ("Non-invasive in-vivo assessment of local stiffness of human artery walls", see [7]) was to fill in the gap with a new proposed methodology. Ideally, the method would rely on measuring the arterial displacements using the ultrasound system by a physician. The measured displacements would be then processed on-the-fly and fed into an inverse solver that would estimate the arterial stiffness and other related quantities of interest. The solver, which most often is an optimization procedure or a Markov Chain [8] requires a lot of evaluations to minimize the error function with respect to estimated parameters and the measured/simulated data. Thus it would need to rely on a fast-to-evaluate model, most likely a statistical one, as using classical numerical models would be too time consuming. Such model would, ideally, be optimal among other candidates and would be based on data generated from a physical (most likely numerical) model of an actual artery. The need for such a numerical model is straightforward: when it is validated on experimental data it can 'replace' the experiments, which are often very expensive and time-consuming to carry out. To fully validate the model, the experiment itself would also need to be assessed in terms of its reproducibility. Furthermore, the experiment would also itself serve as a validation of the

use of ultrasound to reliably read off the displacements produced by an arterial phantom. This work lays the groundwork for the above methodology, by implementing each of the steps and testing them on an experimental phantom.

As mentioned, the main objective behind this work is the development of a non-invasive methodology for local estimation of arterial stiffness. To achieve this, a set of smaller steps is taken. These can be stated as a list to make them more digestible:

- Development of an experimental rig to validate the ultrasound as a valid technique for measuring arterial displacements
- Assessing the quality and reproducibility of the conducted experiments experiments
- Development of a numerical model for simulation of arterial phantoms and its validation
- Surrogate Model development for fast and accurate simulations
- Uncertainty Quantification and Sensitivity Analysis of the FSI model using the surrogate
- Inverse estimation of parameters; particularly arterial stiffness

Let's start from the top of the list. It is a nice idea to use the ultrasound to read off arterial displacements, but it needs to be confirmed experimentally. To that end, first the experimental rig needs to be constructed and the experiments planned. The rig will use two high-fidelity cameras to capture displacements in two orthogonal directions. The used phantom is assumed to be relatively symmetric, thus if the processed data from two cameras is very similar to each other (within error bounds) the symmetry assumption will be confirmed. Then one of the cameras (the top one) will be swapped with an ultrasound and the results compared. To make any of that happen, first appropriate algorithms need to be written that process the data coming from the cameras. The rig will also include pressure and flow meters that allow one to measure the flow conditions in the system and then correlate them with the measured displacements. This allows to the use the data for camera-ultrasoud validation as well as for Boundary and Initial Condition generation.

To make sure the data are not generated by random chance and can actually be reproduced, they need to be statistically assessed: here we move to the second item on the list. The assessment will proceed first by performing a preliminary

statistical analysis on the data to get a 'feel' for it. After this initial processing, a model structure needs to be established that closely matches the experimental procedure. Once such a structure is established, an appropriate model needs to be chosen. When it is chosen, it needs to be compared with the one resulting from the initial, basic preprocessing to prove that it actually is an improvement when it comes to explaining the variance present in the experiment. Once this is achieved, different models of various complexities shall be developed to capture as much of the data as possible and the best one shall be chosen. Subsequently all the cases need to be analyzed and the results need to be checked to see whether they do not invalidate the model's initial assumptions.

Having assessed the uncertainty present in the data, it is time to develop the Fluid-Structure Interaction model. The inputs for it will come from processed experimental data and processed material testing data (these experiments were performed by an external party). Then, two FSI models will be developed: one using the partitioned FSI approach embedded in the ANSYS package and one using the monolithic FSI approach present in the FEBIO package. Mesh sensitivity needs to be performed and models compared against each other. Given that they produce similar results, the faster model needs to be chosen, validated against experimental data and automated for the purposes of developing the surrogate model.

For construction of the surrogate, which is the next item on the list, one needs to choose variables of interest. Subsequently, an experimental design needs to be performed by choosing an appropriate design or sampling scheme. Then, the FSI solver needs to be embedded in a loop and evaluated a prescribed amount of times. The results from these evaluations need to be checked, whether they make physical sense. Then the results needs to be split into train and test sets to allow for surrogate training and then predictive performance checks. Various surrogates need to be trained, their sensitivity to input parameters assessed and ultimately their predictive performance compared against each other. The winning surrogate will be used for subsequent tasks.

The penultimate entry on the list is the UQSA of the FSI model. To do that, first the appropriate method needs to be chosen. Having chosen, it needs to be implemented and checked against the benchmark (e.g. Ishigami function). Correlated parameters, if they exist, need to be treated specially. The convergence of SA must be performed. From UQSA results one is now able to suggest future directions of research when it comes to input parameters, quantify the uncertainty present in the FSI model due to input parameters, and simplify the

model of interest.

Finally, with the surrogate model, it is possible to estimate the parameters of interest (e.g. arterial stiffness). The overarching methodology for the estimation needs to be chosen. Subsequently, the estimation structure needs to be proposed, developed, and the surrogate model embedded within it. The results need to be checked for whether they do not defy the original assumptions of the routine. The final results need to be discussed.

1.3 Fluid-Structure Interaction modeling

In cardiovascular biomechanics, where the solid body (artery or other) enclosing the flowing blood is acted on by it, one needs to consider a joint response of these both physical domains. Classically for separate numerical modeling of fluid and solid domains, two methodologies have been developed; Computational Fluid Dynamics (CFD) and Finite Element Analysis (FEA). FEA is a general framework for analyzing solid dynamics subject to various constraints and forces, employing the Finite Element Method to discretize the Cauchy equations of motion over some arbitrary solid domain, Ω_s, into a linear system of equations, which are then numerically solved to obtain approximations of the considered system [9]. On the other hand, CFD often applies the Finite Volume Method to discretize the Navier-Stokes equations over an arbitrary fluid domain, Ω_f, into yet another system of linear equations [10], which is then solved. While both methods have found wide applicability in engineering and science, they are not sufficient to accurately model cardiovascular mechanics separately [11]. Thus a coupling between these two methods is introduced, termed the Partitioned Fluid-Structure Interaction. It is a method of simulating behaviour of coupled fluid-solid systems. This is often achieved by enabling close communication between the solid and fluid domains during problem setup and solution.

The implementation of Fluid-Structure Interaction can be broadly divided into two approaches: monolithic and ,as desrbied above, partitioned. Within the monolithic approach, the fluid and solid domains are discretized together (usually using the Finite Element Method) and subsequently solved together in one, large linear system. This approach boasts very high accuracy while suffering from very large computational costs. One of the software used for this work, FEBio [12], employs such a monolithic approach. The partitioned approach, on the other hand, works a bit differently. Here, both separate discretization schemes and solvers are used for fluid and solid domains. Information transfer,

instead of being ingrained in the model, like in the monolithic approach, happens at the coupling boundaries where physcial quantities like (for solid-fluid couplings) forces and displacements are exchanged. Within a given time step, convergence of the physical quantities is monitored, which when met allows to proceed to the next time step upon completion of remeshing of both domains. [13] This approach leverages the, separately developed, advanced CFD and FEA algorithms for problem-solving, however it is prone to errors and instability [11].

For the partitioned approach there further exists a distinct difference in domain coupling strategies. When a transfer occurs only from one domain to the other, e.g. pressures calculated within the fluid domain are transferred as forces for the solid domain, one is working with One-Way FSI [14]. It is often applied in cases when the coupling existing between the fluid and solid domains is not very strong and the deformations of the solid body are very low to negligible during the process [14]. Hence it is utilized to primarily estimate the stresses acting on the solid due to the flow of the fluid. Thus One-Way coupling has found many applications, particularly in durability tests of mechanical equipment, i.e. marine propellers [15] or tests on butterfly valves [16], while also having a limited success in medical applications [17, 18, 19, 20], However, when there is a significant modification of the flow field due to non-negligible deformations of the solid element that is either constraining the fluid domain or lies in its path, one should use the Two-Way FSI approach. [14] Here, forces are passed between both domains at the coupled surface(s) and the solution of the domain proceeds as described above. Two-Way FSI has the advantage of producing more realistic results at the cost of increased computational cost due to constant remeshing of the domains and coupling iterations between two domains [14, 11]. Unlike One-Way FSI, it has found wider but still limited applications in biomedical research [19, 21, 22, 23, 24]. Other areas of implementation include ship design and modeling [25], ocean engineering [26] and many others [27, 28, 29]. Recently more studies are being conducted regarding the effectiveness of One-Way FSi vs Two-Way FSI but the conclusion is somewhat repeated: while One-Way FSI tends to compute faster in comparison to Two-Way (usually hours for the former and days, weeks for the latter), it is unrealistic for heavily coupled flows. In a study conducted by Khe [30], it is demonstrated that for coupled flow–vessel interactions, the absolute values of mechanical parameters (such as pressure, stress, strain) tend to decrease. This results in a solution of the hydroelastic problem, which is more detailed and more reflective of the experiment. Thus it seems that the most reliable approach to model the be-

haviour of an arterial phantom, would be to resort either to a monolithic FSI approach or a partitioned one, with Two-Way Coupling.

1.4 Surrogate Modeling

The computational demand of the FSI solvers (both monolithic and partitioned) described above is a big limitation, when one desires to apply the model to something more than just one-off prediction. It is indeed problematic because modern scientific and engineering analyses often demand far more than just singular model evaluation to test some theory. Often many engineering applications such as shape optimization [31], parameter estimation [8], model tuning [32], Uncertainty Quantification [33] or Sensitivity Analysis [34] require that the model be evaluated a large number of times in order to obtain accurate solutions. These approaches often need hundreds or hundreds of thousands of evaluations to obtain a precise estimate, which seems impossible given that for fairly uncomplicated problems and geometries, FSI calculations may take from many hours to many days, weeks. It is thus obvious that in engineering and medicine in particular, days or weeks may be the last straw between a construction failure or patient's death. Thus the limitation imposed by FSI in medical context is a particularly severe one. However modern problems require modern solutions - similarly to how one is able to 'replace' the experiment by developing a numerical model, testing it and then using it to test for new (not experimented) scenarios, it is also possible to develop a second, less involved, statistical model (often called a surrogate) that will 'replace' the full FSI solver. Such a model will naturally possess less resolution than a full order FSI model, the same way an FSI model will possess less resolution than an actual experiment. However using such a surrogate model will drastically accelerate the evaluation process at a loss of a fraction of accuracy. It is often an acceptable trade-off and thus in the engineering and sciences they have been widely applied.

To create such a surrogate model, a more rigorous path is often followed. The surrogate modeling framework can classically be divided into a four steps, often followed (not necessarily rigorously) in this order: [35]

- Design of Experiments
- Variable Selection
- Creating a (surrogate) model

- Model evaluation and selection

Each of these steps appears throughout this work (either explicitly or not) and deserves attention. To have a better idea of methods used in contemporary engineering, science and medicine that encompass either CFD, FEA or FSI modeling on their own, a scoping review has been conducted. Its primary goal was to inform and guide the creator of the thesis when performing and designing analyses. It was carried out according to the procedure described in [36] and its results are summarized in Figure 1.1.

Starting of with the first item on the list above, Design of Experiments (often called Experimental design), is a methodology aimed at reducing the number of experimental evaluations, such as physical experiments or simulations, required to obtain specific information about a process. Historically, Factorial and Fractional Factorial designs were utilized, however, the less restrictive characteristics of computer experiments have led to a preference for space-filling designs like Latin-Hypercube Sampling and Uniform Sampling [37, 38]. This should not be interpreted to mean that Factorial designs have been completely supplanted by space-filling designs; instead, they are predominantly employed in physical experiments rather than in numerical simulations [39]. This result is corroborated by the summarized results in Figure 1.1.

Looking at the top-left part plot in Figure 1.1, it is visible that the Latin-Hypercube Sampling (LHS) method for experimental design is by far the most widely used in the reviewed literature, followed by Uniform and Sequence-based sampling (all being space-filling designs). LHS is by far the most widely applied technique due to its ease of application, straightforward interpretation, and effectiveness [40, 41]. The second most applied scheme is the Uniform Sampling [37, 39, 42], followed by random sequence-based sampling (e.g., Hammersley, Sobol, Halton, Korbov) [38, 43] or Orthogonal Latin Hypercube [44].

A less applied technique that is gaining popularity (especially in the optimization community) is variational (sequential) sampling (particularly prevalent in the literature on Gaussian processes) [41, 45, 46]. There are other techniques, far less present in the chosen literature, like Linear Nearest Neighbor [41], greedy sampling on the output (or both on inputs and outputs) [47], Spatial Simulated Annealing (particularly for working with spatially related data) [48], Custom Sampling [49]. However, in engineering applications, it is also relatively common to use domain knowledge to limit possible model evaluations and limit explorations to a few design points [37, 50].

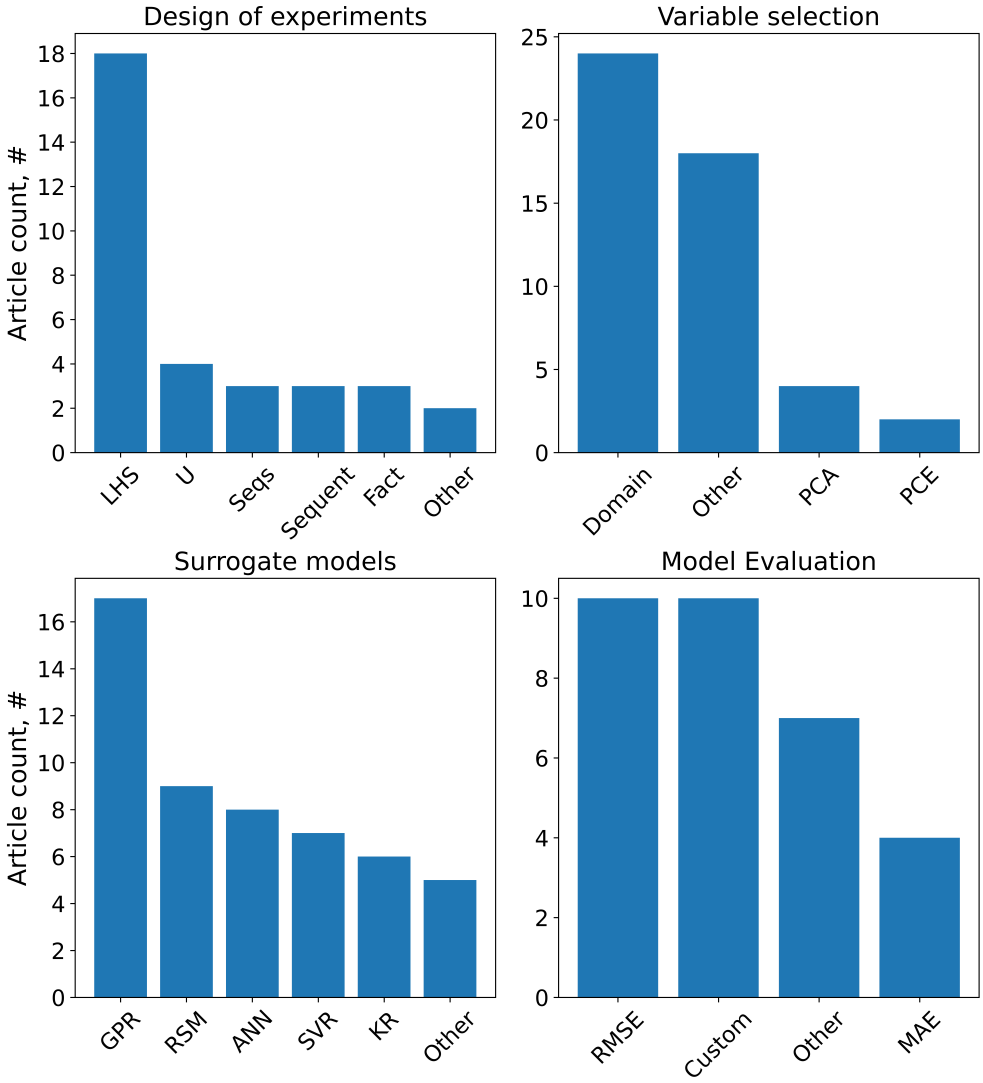

Figure 1.1: Summary of results of the literature review. Each of the four graphs shows the frequency (prevalence) of appearance of methods in the reviewed papers. Shortcuts stand for: LHS - Latin Hypercube Sampling, U - Uniform Sampling, Seqs - Sampling based on Random Sequences, Sequent - Sequential Sampling, Fact - Factorial Design, Domain - physical or domain knowledge, PCA - Principal Component Analysis, PCE - Polynomial Chaos Expansion, GPR - Gaussian Process Regression, RSM - Response Surface Modeling/Methodology, SVR - Support Vector Regression, ANN - Artificial Neural Networks, KR - Kernel Regression, RMSE - Root Mean-Squared Error, MAE - Mean Absolute Error

The second plot, top-right, in Figure 1.1 represents the selection of variables (first item on the list), that is, which variables are deemed "important enough" for a surrogate model. Just as in experimental design, it is commonly predetermined in engineering which variables are crucial for the modeling framework. Often these variables are related either to the physical process that is being emulated [37, 39, 45, 49],[51], [52], to the geometry of the object [40, 52, 53]. They might also embody a parameterization of this geometry [42]. Variable selection is frequently closely related to experimental design, as the techniques employed to develop sampling strategies also yield insights that help distinguish critical variables. For example, employing fractional factor designs can help reduce the set of significant variables [52].

Often, the variables in question are either too abundant or fall outside of conventional understanding. Under these circumstances, it is typical to utilize dimensionality reduction methods such as Principal Component Analysis [38, 54, 55] or to express the data in a different (usually orthogonal) framework, such as through Polynomial Chaos Expansion (PCE) [56].

The third panel represents the surrogate model itself (third item on the list), that is, the emulator chosen by the researchers for the given task. Gaussian Process Regression (GPR, or more generally Gaussian Processes, GPs) has been frequently selected as the go-to method, describing its flexibility, predictive capabilities, and mathematical tractability as the primary selling points [55, 57]. GPs are also applied due to their statistical setting and providing readily available and easy-to-understand uncertainty estimations [45]. The most prevalent form of GPR features a zero-mean function [55] and a Squared Exponential kernel (SE) [38, 43] [55] - [58]. However, researchers often employ non-SE kernels or construct more sophisticated kernels (e.g., the combination of inputs of linear order with interactions and SE kernel in [38] or polynomial combinations [58]). Furthermore, some papers strayed from the pack in terms of mean functions, making them nonzero in order to capture some trend [59, 60]. This was notably interesting in [59] where a linear mean function was selected to mimic the behavior of a hyperelastic material model. Alternatively, some researchers apply Gaussian Processes within the Kriging framework solely to characterize the noise component of a model [49] - [60].

Often Gaussian Process models are benchmarked against other, more standard and well-described models such as Response Surface Models (Polynomial Regression) or Artificial Neural Networks [47] or simply used as an explored (but not taken) path in surrogate modeling [49]. Gaussian Processes are also widely

used in problems related to numerical model calibration (or parameter estimation or inverse problems) using the Bayesian Optimization framework, particularly with the Expected Improvement method [58]. When the data fundamentally do not follow a Gaussian distribution and could be better represented by a different distribution, an alternative method can be employed. When the data deviate from a Gaussian distribution, relying on existing analytical formulas to train the GP model becomes impractical. Consequently, it becomes necessary to employ optimization methods that utilize non-Gaussian likelihoods. These methods, although they enhance the predictive capacity of the model, require greater computational resources [59].

Another highly favored approach, particularly prominent in data science, that is increasingly adopted in engineering disciplines is Artificial Neural Networks (ANN). Currently, they are often called Deep Neural Networks, a name derived from their use of numerous layers populated by a large number of neurons [39, 50, 49, 61]. When the interpretability of the model (as in cases like Linear Regression or SVM) or the statistical analysis of the outcomes (as seen in GPR) is not a priority, simple ANNs prove to be effective in prediction [41]. Due to the large number of model parameters and non-linear activation functions, they can often approximate complex phenomena. The most common configurations apply sigmoid [37, 50], hyperbolic tangent (tanh) [39], linear [50] or ReLU [61] as layer activation functions. Since ANNs, like GPs, have begun to gain widespread adoption only in the late 1990s, it is fairly common to compare them with more traditional and well-established models such as linear or polynomial regression (generally referred to as RSM). [39, 50].

One of the tried and tested techniques is the Response Surface Methodology, which relates given response variables to a set of combinations of predictor variables, usually representing powers of variables or variable-crossed products up to a given power. Knowing the possible structure of the relationship that produces the response, RSM has historically been widely used as the primary surrogate model or a reference model [39, 40, 49, 62]. Beyond Gaussian Processes, Artificial Neural Networks, and Response Surface Methodology, various other well-known techniques exist that, although not highlighted extensively in this discussion, are still mentioned and sometimes employed or evaluated. Some of these methods are Random Forest Regression (also Decision Trees) [47], Polynomial Chaos Expansions (particularly in Uncertainty Quantification and Sensitivity Analysis, UQSA, context) [54, 63], Genetic Algorithms (particularly in the Mulit-Objective Optimization setting) [60, 64], Support Vector Machines

[41, 43, 53] (very often employing a kernel), kernel-based regression methods (most often using the Radial Basis Function kernel) [42, 65], Radial Basis Functional Networks [66],K-Nearest Neighbors [61] and others. Surrogate models are frequently integrated into what are termed 'ensemble' models, forming a 'supermodel' that enhances predictive accuracy while preserving the adaptability of its individual components [51, 62]. Such an ensemble is presented in [51] where a combined model consisting of Gaussian Processes, Support Vector Regression, and Polynomial Chaos Expansions is created. Combining one or more neural networks in a similar 'ensemble' is also possible [50]. Furthermore, there is a notable trend towards the adoption of gradient-enhanced surrogate models, leading to higher precision in surrogate models[43]. One of the interesting applications of surrogate models is to train them to project the original data onto some reduced basis obtained from a dimensionality order reduction process (e.g. creating a Gaussian Process to project onto Principal Components [38]) to emulate and speed-up the order reduction process itself. It is also common to project from the reduced basis to the original results (e.g. Principal Component Regression, often uses an RSM surrogate) [48].

Once experiments have been designed and the surrogate model has been properly fitted, assessing its predictive performance (final item on the list) is crucial. This is due to the fact, that if the primary goal of the model is to predict unobserved responses of the system, it must be good at it. Another common tactic, resulting from the model performance, is to compare the obtained model's performance against several other ones. This is done in order to determine the most suitable model before committing to one or the other. To assess model performance, one must look at performance metrics, also termed error metrics. They can be classified by error type or by particular metric. To define the former, the two most common phrases in machine learning are 'train' and 'test' (or 'validation') error. The first describes the discrepancy between the model's predictions and the training data on which it was developed. Testing error, on the other hand, refers to the inaccuracies observed when the model predicts outcomes for previously unobserved data. The literature defines many common metrics for quantifying the error. The most prevalent metrics are the Mean-Squared Error (MSE, or its counterpart the Root Mean-Squared Error, RMSE) [41] - [45],[48]-[56], [65]. Metrics based on absolute differences are also used, particularly when there is a suspicion of outliers, as absolute errors are less prone to them [47, 49, 63]. Relative errors are employed when one wants to compare the computed error (be it squared, absolute etc.) to the 'ground truth values' to better indicate the magnitude of the model's departure from it

[41, 47] Other metrics are also used [46, 51, 64]. Among other techniques used, one can find Negative Log Predictive Density (NLPD) [38], Maximum Error [44], Squared Sum of Errors (SSE) [58], Fractional Bias [55] and others. It is important to note that many articles often report the R^2 value (also called the 'coefficient of determination') [50]-[55],[59], however it is only a goodness-of-fit measure (indicates how much of the data variance is explained by the model) and not a predictive assessment metric. [67]

The prevalence of surrogate models can only be corroborated by the variety of domains of their application. They range from non-linear material modeling [59], fast parameter estimation of biomechanical material properties from clinical and experimental data [61], determining device- and species-specific hemolysis power law coefficients [45], urban atmospheric flows modeling [55], structural stability and damage of engineering structures [40, 56, 57], design exploration and shape optimization [43] - [49], [60, 64], structural [51] or soil [54] reliability, full order model calibration [58], predicting possible failure in manufacture processes [57], predicting fine flow-field details based on a robust model trained on low-fidelity data [50], UQSA of Models (e.g. flow around a ship in shallow water) [63] and many others. Over the last 20 years (approximate time span considered in the review), there has been a wide variety of surrogate modeling approaches used in Finite Element Analysis and Computational Fluid Dynamics. Each of the four paradigms of surrogate modeling [35] has been extensively used in papers, each having a clear favorite method. For the design of experiments, LHS is the go-to method due to its simplicity and space-covering properties. For variable selection, however, statistical approaches have been replaced by experts' domain knowledge. This makes sense in the domain of engineering and sciences, as often there is a clear physical law with identified inputs and outputs, that underlie the analysis. For the selection of the surrogate model itself, Gaussian Process Regression is the clear winner. Its flexibility, accessibility, and interpretability make it the most widely applied method. For model performance evaluation, the most popular choices are RMSE or custom. The enduring popularity of RMSE, given its long-standing and proven track record, contrasts with the equally favored custom methods. It is logical that with the increasing complexity and specialization of topics, priorities shift, rendering basic metrics such as RMSE and MAE insufficient for the authors' requirements. Additionally, it is important to acknowledge the wide range of disciplines that employ CFD, FEA and FSI (although these are few and far between) along with the SM framework. It ranges from material modeling and biomechanics to structural stability or urban atmospheric flows.

This proves that the Surrogate Modeling framework can be applied almost anywhere, which makes it so popular among scientists and engineers.

1.5 Uncertainty Quantification & Sensitivity Analysis

In reality each parameter that constitutes a numerical model built on top of physical principles (like fluid viscosity, material stiffness, pressure) or a statistical one (built solely on data, e.g. dependence on carotid artery stiffness on age, sex, ethnicity etc. [68]) is subject to some uncertainty. In general the total uncertainty of a system is commonly defined to be built of two constitutents: aleatoric uncertainty and epistemic uncertainty. [69] Aleatoric uncertainty refers to the one, which has an inherently random origin. An example may be equipment noise present in measurements. At some point it is something that generally cannot be reduced any further, and remains in the system's measurement. This it the reason that aleatoric uncertainty is often termed 'irreducible' uncertainty. Epistemic uncertainty is the one resulting from experimenter's lack of knowledge and generally can be reduced with more information. A simple example is measuring physical distances. The uncertainty regarding the accuracy of the measurement can be driven down by using more and more refined tools, e.g. going from 'eyeballing' dimensions to using a laser rangefinder. Thus an alternative name for epistemic uncertainty is the 'reducible' uncertainty. [69] Thus typically the analyst is concerned with both types of uncertainty: the irreducible part is usually quantified and theh kept in the back of the mind as the inherent randomness of the system. The reducible part is quantified in hope of reducing it to the smallest value possible. It is important to note, that the division of the total uncertainty in the model, between the aleatoric and epistemic one is not known beforehand. This is generally what Uncertainty Quantification is, and it is most often done by calculating statistical moments with respect to the model output. What allows one to reduce the epistemic uncertainty is variance-based Sensitivity Analysis, SA (as opposed to other methods like e.g. derivative-based methods, which are not used in this work). [70]

SA broadly refers to the practice of investigation of the contribution of various changes in model assumptions, parameters, variables etc. to the final model output resulting in uncertainty. In surrogate modeling and statistical modeling, it primarily refers to the amount of variance stored in input parameters and

variables and how much of an influence they have on the final output variance [71]. Due to the fact that the parameters and variables may appear in various relations (e.g. quadratics, logarithms) it is not so straightforward to claim that a given parameter is the most influential simply because it has the largest variance [72]. It is even more murky in situations where there is no analytical or correlation-based expression for model response, which is most often the case. Thus to be able to systematically carry out SA tasks, many approaches have been developed. The primary categorization is the division into global and local Sensitivity Analysis. Local sensitivity analysis focuses on providing in-depth insights on how the parameters of a model affect its inputs at specific points within the design space (such as through derivative-based methods) [73]. Although this method is attractive, it only measures the sensitivity within a narrow local area, which can result in overlooking significant areas, especially in non-linear and complicated models (or models when the structure is wholly unknown) [72]. On the other hand, Global Sensitivity Analysis aims to capture the general sensitivity of the model resulting from its inputs by distributing the contribution of individual parameters (and their interactions) to the total output variance [74]. This method inherently benefits from its ability to identify global patterns in the model by generating an importance hierarchy for the overall variable. In summary, the combination of UQ and global variance-based SA allows one to quantify the total amount of variance present in the model and apportion this variance among the inputs and parameters of interest. Subsequently, SA also advises one where to take the analysis further. First, it points one in the direction of non-influential inputs that can be fixed at their reference (literature, experimental, etc.) values disregarding their randomness. This allows one to simplify the model under consideration and either wholly disregard certain parameters or keep them fixed. Secondly it directs one to the least 'explored' parameter, one whose further investigation and repeated measurements, would allow to reduce the total output variance the most. [72]. Thus UQSA is a very useful tool for anyone developing models and looking to gain insight from them.

1.6 Inverse Problems

It is often of interest to engineers and scientists to infer parameters present in physical system. However it is not always easy to measure quantities like the modulus of elasticity or thermal conductivity. However it is easier to measure things they produce, e.g. amount of heat as a result of temperature differential

or the stress produced as a result of imposed strain. Then the next thing to do, is to estimate this physical property from the simple measurement. This is generally called an Inverse Problem [8]. More generally, Inverse problems are mathematical and computational approaches where the goal is to infer the causes or parameters of a system under consideration, based on observations of its effects or output. In other words, in an inverse problem, one seeks to determine the inputs or parameters of a system that are consistent with observed outputs or measurements. These problems are prevalent across various scientific disciplines and engineering fields, including physics, engineering, geosciences, medicine, finance, and many others. [8] The distinguishing feature of inverse problems is their inherent ambiguity or ill-posedness, which arises due to several factors: [8]

- Non-Uniqueness: There may be multiple solutions or sets of parameters that could produce the observed data, leading to non-uniqueness in the solution.
- Instability: Small errors or noise in the observed data can lead to significant uncertainties or errors in the inferred parameters, making the solution sensitive to measurement errors.
- Ill-Conditioning: The relationship between inputs and outputs of the system may be ill-conditioned, meaning that small changes in the inputs can result in large changes in the outputs, exacerbating the sensitivity to noise or errors.

These challenges lead to the development of various methods for conducting inverse analyses and interpreting the results. Inverse problems can be broadly categorized into two types based on the nature of the relationship between inputs and outputs. The first category belongs to Deterministic Inverse Problem. There, the relationship between inputs and outputs is governed by deterministic mathematical equations or models. Examples include parameter estimation in ordinary and partial differential equations, system identification, and inverse modeling in engineering and physical sciences. [75] The employed methods include (but are not limited to) Ordinary Least Squares (OLS), Tikhonov Regularization for OLS, Levenberg-Marquardt, Gauss-Newton method or Genetic Algorithms. [8]

In statistical inverse problems, the relationship between inputs and outputs is characterized by stochastic processes or random variables. The statistical approaches have a natural upperhand over the deterministic ones in the sense of

providing some quantification of uncertainty related to the estimated quantities and/or possibly other things like the estimation procedure itself. Many methods for statistical inverse problems are applied e.g.: Maximum Likelihood Estimation, Generalized Method of Moments (and more generally frequentist methods), bootstrapping techniques or Bayesian inference. [8] Due to its efficient and robust way of quantifying uncertainty, Bayesian inference has found wide applicability in the field of medicine, particularly, medical computer vision tasks, material parameter estimation, disease modeling, Bayesian networks and more [76]. In this work the statistical forumation is sought in order to not only provide an estimate for artertial stiffness, but also simultaneously provide uncertainty related with it, that can facilitate the decision making process on the physician's part. [77]

1.7 Thesis Outline

The thesis is divided into six main chapters, all revolving around the development of a surrogate model useful for fast prediction of displacements, which can be embedded into problems that require many evaluations, that are costly for a regular FSI model. The Thesis begins with an *Introduction* which provides an overview of the main topics contained within the thesis as well as results of a scoping literature review that was conducted on surrogate modeling in contemporary engineering and scientific literature. The results of the literature review served as guidepost for the researcher while conducting the analyses present in the latter half of the thesis.

Second chapter, *Mathematical and statistical methods*, introduces methods that facilitated the experimental data processing (e.g. image processing filters, Hampel filter for outlier removal), statistical theoretical background to facilitate experimental data analysis and modeling (e.g. statistical inference, ANOVA, Linear Mixed-Effects Modeling) and finishes with the introduction of Governing Equations that are used in the formulation of the Fluid-Structure Interaction methodology.

Consequently, the third chapter, *Surrogates, UQSA and Bayesian Inference* focuses on the theory behind the surrogate modeling methodology (e.g. experimental design, model selection), describes the two models considered in the work for FSI model emulation (Gaussian Process Regression, Data-driven Reduced Order Modeling). Further it develops concepts needed to understand and perform Uncertainty Quantification and Sensitivity Analysis of the FSI model

with the aid of the surrogate. Finally, it introduces Bayesian statistics and how it can be used to inversely estimate the parameters of interest, e.g. the arterial stiffness.

Next is the fourth chapter, *Statistical analysis and FSI validation.* It starts with the overview and operation of the experimental rig with an arterial phantom. Then the pressure, flow and image data processing is detailed. What follows is the description and development of the Linear-Mixed Effects models that were fitted to experimental data in order to assess the data's (and by extension experiments) reproducibility and accuracy. A simpler statistical analysis is applied to material testing data that was obtained from an external party. Subsequently the flow, pressure, image (displacements) and material testing data are used in FSI model development and the results from conducted statistical analyses are used to validate its performance.

The penultimate chapter, *Surrogate development and application* describes how the development of the surrogate model was effected and its applications. It begins by first describing the design of the experimental space, which hinges on results from the previous chapter and literature data. Next is the selection of the appropriate sampling algorithm. Then the FSI model is automated and evaluated N amount of times, the results are discussed and compared with literature data. What follows is the fitting of two models to this data. The models considered in this work are: an SVD-based Reduced Order Model and (Sparse) Gaussian Process Regression. The sensitivity of these models to corresponding hyperparameters is explored and the better performing model selected for the upcoming tasks. Next is the implementation of the Variance-Based SA through Monte Carlo estimators of the Sobol Indices. Three various methods are implemented by the author and compared against analytical literature benchmark data and against another method (Polynomial Chaos Expansion). The results are discussed and model is tuned accordingly. The last application of the model is the phantom stiffness estimation. The Gaussian Process surrogate is embedded within a Hierarchical Bayesian framework and is used to estimate parameter values at different levels (subject and group levels). The results of the estimation, and the quality of the estimatation process itself, are assessed and discussed.

The thesis finishes with the *Conclusions* drawn from the conducted analyses and proposes many avenues for further research that could extend the methods and topics applied within it.

2

Mathematical and statistical methods

In order to validate any model, be it numerical or statistical, its primary goal is to be able to recreate some conditions of interest. [78] This often tends to be a long and arduous process that begins in the laboratory and ends with an analyst confirming that the model results mimic the physcial, experimental ones. The end goal of Chapter 4 is the assessment of reproducibility of performed experiments and the validation of a FSI model on the experimental data using the calculated data uncertainty. Thus the contents of this Chapter consitute the theory of all the applied methods to achieve the aforementioned goal. It begins with the basics of digital images and their processing as well as the introduction of an outlier identifying filter for pressure and flow data. It then proceeds to lay the groundwork for the method used in the statistical assessment of experimental data - Linear Mixed-Effects Models. To do this, the concepts of Statistical Inference, ANOVA, random and fixed effects need to be fleshed out first. What follows is a brief discussion on the benefits of applying Linear-Mixed Effects models to model experimental data. Finally the chapter wraps up with the introduction of Governing Equations of the Fluid Structure Interaction model.

2.1 Images and Basic Image Processing Techniques

Usually when one thinks of an image, a photograph comes to mind (whether taken with a professional camera or a phone). This implies that the image under consideration is of the digital type. A basic building unit of a picture is a pixel, which is a single value, usually between 0 and 255 (i.e. an 8-bit representation) [79]. The image's resolution is directly related to its pixel count. For a grayscale image it represents the intensity of the image, 0 refers to an all black image, while 255 refers to an all white picture. However usually one first deals with color pictures. Such images leverage the simple 0-255 representation, by using different arrays of pixels called channels. The most common channel composition in usage is RGB - Red, Green and Blue. A color image is composed of a sum of these channels; see Figure 2.1. Other channel representations exist such as BGR (Blue Green Red - just a change in color ordering) or HSV (Hue, Saturation, Value - this is an entirely different way of representing an image).

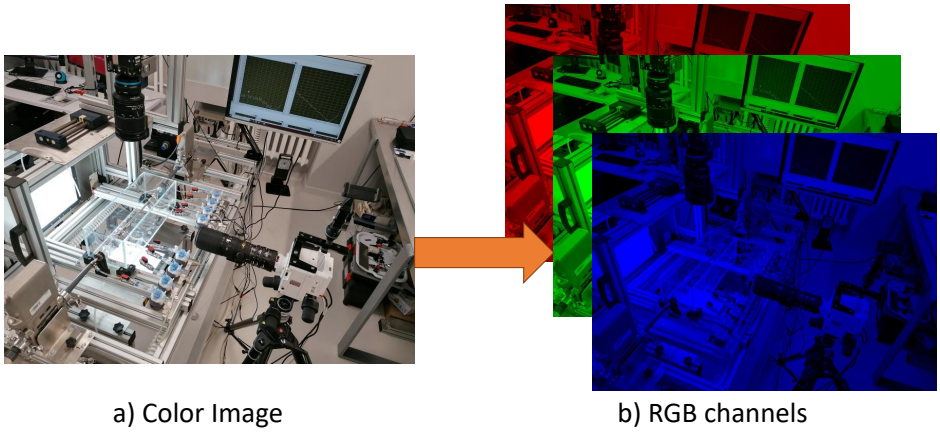

Figure 2.1: Decomposition of a color image into its three components: the Red, Green and Blue (RGB) channels.

One of the most common types of operations to perform on an image is linear filtering. It is a neighborhood-based operation (i.e. it uses a collection of pixels near to a pixel to determine its output value). The linearity in the name comes from the fact, that to produce new pixels it uses a linear weighted combination of pixels that are close by. It can be mathematically written down as: [80]

$$g(i,j) = \Sigma_{k,l} I(i+k, j+l) h(k,l) \tag{2.1}$$

where I is the original 2D image and $h(k,l)$ is the so-called filter mask, whose entries (weights) are called filter coefficients. This operation is called cross-correlation and may be written as:

$$g = I \otimes h \tag{2.2}$$

However upon reversing the signs in I, one obtains an operation of convolution, commonly written as:

$$g = I * h \tag{2.3}$$

Unlike convolution, cross-correlation produces a reflected version of an original image, and thus is used more widely for basic image processing tasks. [80] The convolution operation is visualized in Figure 2.2. As can be seen, a 3×3 filter is passed through an image, producing an output. A common problem in filtering is that images shrink, thus they need to be padded at the edges to avoid this phenomenon. There is a wide range of various filters available for use with a wide range of results that they produce.

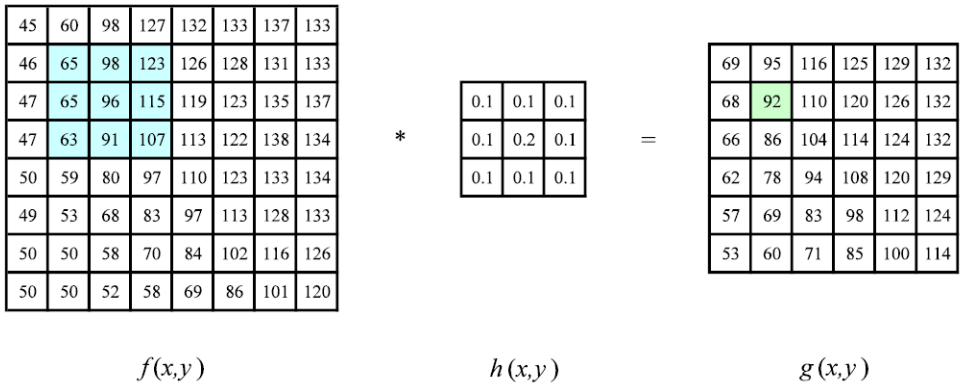

Figure 2.2: Convolution operation on an image f(x,y). Reproduced from [80]

Gaussian Filter is used to blur images. It has the mathematical from similar to an actual Gaussian Distribution (see section 2.3) [80]:

$$h(x, y; \sigma) = \frac{1}{2\sigma^2\pi} exp(-\frac{x^2 + y^2}{2\sigma^2}) \tag{2.4}$$

the key difference between a standard formulation of the Gaussian distribution and the Gaussian kernel is that in the argument of the exponent there is $x^2 + y^2$ instead of $(x - \mu)^2$. x, y stand here for coordinates in the image (pixel values).Gaussian Filters weights add up to 1 and for a 3x3 filter are [81]:

$$h_G = \begin{pmatrix} \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \end{pmatrix} \tag{2.5}$$

an example of a Gaussian filter application can be seen in Figure 2.3 c. Gaussian Filters are often used to de-noise images and prepare them for edge detection. Before applying a Gaussian Filter the image has to be converted to a Grayscale Image (see Figure 2.3 b), which is a common practice in image processing.

A Sobel Filter or Sobel Operator is a mask used to find the differences between neighbouring pixels. It has been widely used in edge detection, often as part of edge detection algorithms. The Sobel Filter is an approximation to a First Order Derivative in a given direction and it can be explicitly stated as [80, 82]:

$$H_{S,x} = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$$
$$H_{S,y} = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \tag{2.6}$$

Sobel Operator can act in X, Y or both directions. Its application to an image in X-direction can be seen in Figure 2.3 d, demonstrating that some edges in horizontal direction are starting to show. Generally, it is common to have the X-Y directions combined to get the overall Gradient, G [80, 82]:

$$|G| = \sqrt{H_{S,x}(I)^2 + H_{S,y}(I)^2} \tag{2.7}$$

It is also possible to find the orientation of an edge [82]:

$$\angle G = arctan(\frac{H_{S,x}(I)}{H_{S,y}(I)}) \tag{2.8}$$

Another useful processing technique is **thresholding**. Depending on the type, it is done by suppressing parts of an image based on some cutoff value (i.e. threshold), e.g. for a binary threshold [83]:

$$I^*(x,y) = maxval, \forall I(x,y) > \text{thresh} \tag{2.9}$$

It is a binarizing operation, that makes parts of the image that satisfy the condition set to some maximum value (e.g. 255), while the rest is 0, resulting in a truly black and white image (see Figure 2.3 e). Often images possess different lighting in different regions, that's where adaptive thresholding is used.

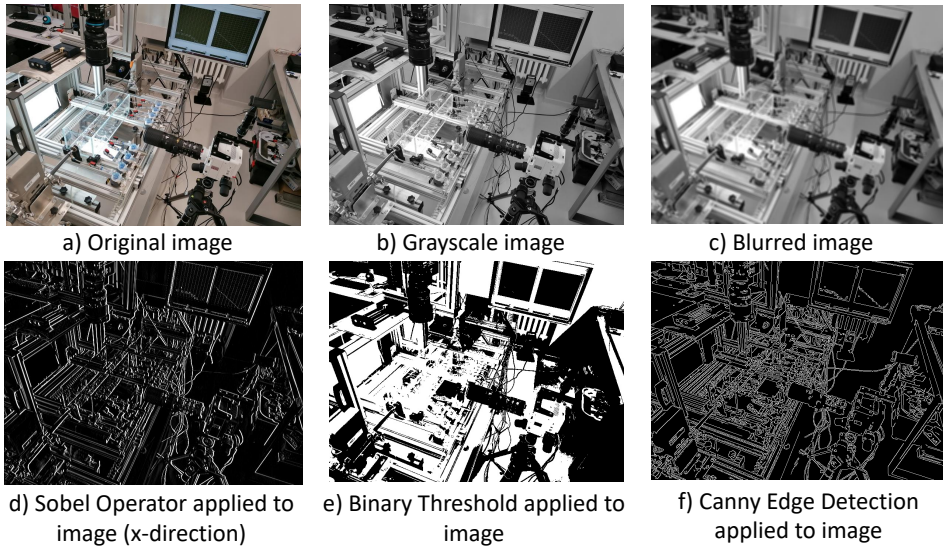

Figure 2.3: Various image operations: **a)** Raw image. **b)** RGB image turned to Grayscale. **c)** Image blurred with a Gaussian Filter. **d)** Sobol operator in a horizontal (X) direction applied to an image. **e)** A binary threshold applied to a picture, with a cutoff of 127 and maximum equal to 255. **f)** Canny Edge Detection applied to an image

One of the fundamental tasks of image processing is edge detection and one of the most popular algorithms used for edge detection is the Canny Edge

Detection. It consists of a series of steps, some of which have already been described and visualized [84, 81]:

- Image conversion to grayscale
- Gaussian Blur
- Determining the intensity gradients using the Sobel operator
- Non-max suppression
- Double Thresholding
- Edge tracking using hysteresis

So far, only the last three steps have not been described. Non-max suppression utilizes the computed gradient magnitude and direction. It proceeds by doing a full scan of the considered image in order to get rid of any pixels that most likely do not make up an edge. To achieve this, each pixel in the direction of the edge is compared to its neighbors (in the same direction), if it is indeed a local maximum, it is considered an edge and the algorithm proceeds to the next example. [81] The next operation, Double thresholding, is centered around selecting two threshold values, let's say $t_1 = 50$ and $t_2 = 100$. In this way, anything that has lower pixel intensity than t_1 is set to 0 (i.e. essentially removed from the image). Values that lie between threshold values are set to be 'weak edges'. Thus it follows that pixels above t_2 are set to be 'strong edges'. With this classification of pixels the last step is Edge Tracking using Hysteresis. In this step an algorithm steps through the image and decides on edges: any weak edges that are not attached to the strong edges are discarded, while ones that are, are kept. The final result of Canny Edge detection applied to images can be seen in Figures 2.3f and 4.4.

2.2 Hampel Filter

All experimental data tend to have some sort of outliers, which can be defined as measurements straying too far from the bulk of the data, thus being an unrepresentative part of it [85]. Thus often the first step of data processing is the removal of any present outliers that may impede proper data analysis. One of the methods appropriate for time-series data is the Hampel Filter, which is based on the idea of using robust statistics to estimate the bulk of the data [86]. One of such robust statistics is the median. Thus, Hampel Filter utilizes

a multiple of the Median Absolute Deviation (MAD) to discriminate against outliers. The purpose of its operation is as follows. First, a window is selected, within which the calculations are performed, i.e. $\langle x_{i-w}, x_{i+w} \rangle$, where x_i is a datapoint and w is the window size. Within the window, a median of the values is calculated:

$$\tilde{X} = \text{median}(\langle x_{i-w}, x_{i+w} \rangle) \tag{2.10}$$

Then, Median Absolute Deviation is calculated for each of the datapoints within the window:

$$MAD = \text{median}(|x_i - \tilde{X}|) \tag{2.11}$$

Basing on the calculated MAD value, a simple signal thresholding (for explanation of thresholding see Section 2.1) principle is employed, where if a datapoint exceeds some multiple, k, of the MAD it is treated as an outlier [86]:

$$\mathbb{I}_{\text{outlier}} = \begin{cases} 0 & |x_i - \tilde{X}| \leq k \times MAD \\ 1 & |x_i - \tilde{X}| > k \times MAD \end{cases} \tag{2.12}$$

where $\mathbb{I}_{\text{outlier}}$ is a binary indicator function. Subsequently if the datapoint is deemed to be an outlier, it can be replaced with some other value, i.e. with the median $\tilde{X}$. A good value for k is often somewhere between 2 and 3 and the width of the window, w, determines how much the local data will influence the detection of the outlier. [86] If still there were some residual high frequency components in the data they were removed using a low-pass filter [87].

2.3 Gaussian distributions and their properties

The crux of the classical statistical inference and modeling (as well as other methods used in this work, e.g. Gaussian Process Regression) is the Gaussian (Normal) distribution. The Univariate (i.e. one dimensional) Gaussian Distribution is defined as [67]:

$$p(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{(x-\mu)^2}{\sigma^2}} \tag{2.13}$$

i.e. it is parameterized by μ, which is the mean and σ^2 which is the variance, the affect the location and width of the Gaussian respectively, the distribution of a variable X is written as $X \sim N(\mu, \sigma)$. A Standard Normal variable, X is defined as $X \sim N(0, 1)$ and obtained by: [67]

$$X^\star = \frac{X - \mu}{\sigma} \tag{2.14}$$

where X is a random variable. It is important to note that this procedure is applicable to any variable. In Figure 2.4 A one can see the probability density function of the Standard Normal and how changing the parameters affects the shape of the distribution. An extension of the univariate normal variable to multiple dimensions (i.e. multiple random variables) is the Multivariate Normal. Some vector-valued random variable $X = [X_1, ..., X_n]^T$ has a multivariate normal distribution when its probability density function is as follows [88]:

$$p(x; \mu, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{\Sigma}|^{\frac{1}{2}}} exp(-\frac{1}{2}(x - \mu)^T \mathbf{\Sigma}^{-1}(x - \mu)) \tag{2.15}$$

here μ is a (mean) vector, $\mu \in \mathrm{R}^n$ and $\mathbf{\Sigma}$ is a covariance matrix, $\mathbf{\Sigma} \in \mathrm{R}^{\mathrm{nxn}}$. This means that a vector-valued random variable, can be thought of as a collection of random variables that are jointly distributed under a joint, multivariate Gaussian distribution. The primary interest in this formulation is the covariance matrix, $\mathbf{\Sigma}$. For a pair of random variables X_i and X_j, the covariance is defined as:

$$Cov[X_i, X_j] = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] - \mathbb{E}[X_i]\,\mathbb{E}[X_j] \tag{2.16}$$

where $\mathbb{E}$ denotes the Expectation [67]. It provides a succinct way to describe the linear dependence of two variables. The covariance matrix is constructed as follows:

$$\Sigma = \begin{pmatrix} X_{1,1} & X_{1,2} & ... & X_{1,j} \\ X_{2,1} & X_{2,2} & ... & X_{2,j} \\ ... & ... & ... & ... \\ X_{i,1} & X_{i,2} & ... & X_{i,j} \end{pmatrix} \tag{2.17}$$

where $X_{i,j}$ has been used to mean $Cov[X_i, X_j]$ to save space. In the covariance matrix, the diagonal elements represent the variance of given random variables,

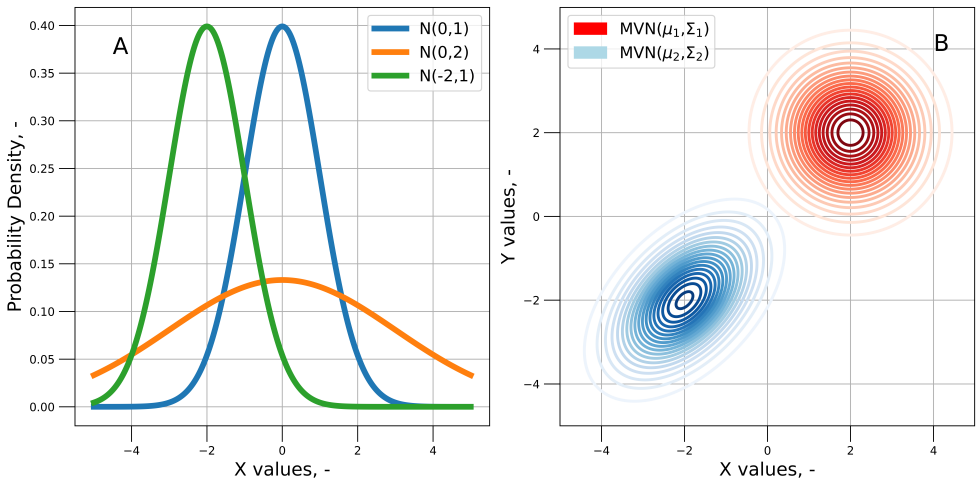

Figure 2.4: Probability densities of Normal and Multivariate Normal Random Variables. **A**: A probability density function of a univariate RV showing the influence of parameters on curve location and shape. **B**: A multivariate (bivariate) normal variable showing the changes in shape with changing mean vector and covariance matrix.

while the off-diagonal terms represent their covariance. A bivariate Gaussian distribution (i.e. jointly defined for two random variables, X and Y) can be seen in Figure 2.4B. Here the change in the mean vector moves the distribution around the (X, Y) space, while increasing the covariance (correlation) between the variables makes the distribution increasingly elliptic reflecting their relationship , i.e. in Figure 2.4B blue contours higher positive correlation, while the red contours have no correlation, i.e. X and Y are independent. To be a covariance matrix, a given matrix must be positive semi-definite, symmetric and invertible (full rank). The symmetry of the covariance matrix follows naturally from the definition of covariance, where $Cov[X_i, X_j] = Cov[X_j, X_i]$. The positive semi-definiteness of the covariance matrix ensures that eigenvalues of the covariance matrix are strictly positive or equal to zero. This can be naturally understood as eigenvalues encode a lot of information about variability, e.g. the variance along Principal Components, meaning they cannot be negative. [89, 90] Finally the invertibility is not a crucial demand for a matrix to be a covariance matrix, however for the purposes of this work (particularly in Gaussian Processes) there is a need to be able to calculate the inverse of the

covariance matrix, Σ^{-1}. One of the properties of primary use in this work is the conditioning in Multivariate Gaussian distribution. By allowing X to be a random vector distributed as a Multivariate Normal variable [90]:

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim MVN(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}) \tag{2.18}$$

where the vector and matrix in the MVN's definition are the mean vector and covariance matrix respectively. Following this set up, the conditional formula is [90]:

$$X_1|X_2 \sim N(\mu_{1|2}, \Sigma_{1|2}) \tag{2.19}$$

where the marginal conditional mean, $\mu_{1|2}$, is calculated as [90]:

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2) \tag{2.20}$$

and the marginal conditional covariance, $\Sigma_{1|2}$, is [90]:

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \tag{2.21}$$

This implies that even when one assumes that the mean value of variable X_1 is 0 it can still come out to be non-zero when conditioned on X_2, due to its dependence on X_2. This result and formulas in equations 2.20 and 2.21 are crucial for the implementation of Gaussian Processes. For the proof of these formulas and more in-depth description, the reader is referred to [90].

2.4 Statistical Inference in Linear Regression

To be able to statistically analyse the experimental data it is crucial to introduce the basic notions of statistical inference. One of the goals of statistical inference is to be able to reach conclusions with an adequate degree of certainty based on a limited amount of data. [67]. In statistics, a clear distinction is made between the population and (random) sample data. The basic assumption in the frequentist statistics is that there exists some true, yet unknown, population. It can be a tangible population, like population of people's heights in Poland, or some less tangible one like error distribution of various models given

some data. Unfortunately one is unable to access the true population and is restricted to relying on finite size samples that are usually a very small subset of the true population. Such a subset is called a random sample and based on this random sample, the goal is to be able to extract as many insights as possible. Often one assumes a given distributional form the data follows, e.g. the Normal distribution, and then tries to estimate its parameters, i.e. the mean μ and standard deviation, σ. This leads to obtaining estimators, commonly denoted with a hat, i.e. $\hat{\mu}$ and $\hat{\sigma}$. When only one value is sought to represent the given parameters, it is called point estimation [67]. Although point estimation is useful it never really tells anything about the range of variation of the calculated values and thus interval estimates were introduced to provide the uncertainty range.

2.4.1 Sampling distributions and Confidence Intervals

Applying the standarization procedure but for the case where the random variable is the estimator of the mean, i.e. $\mu \approx \hat{\mu} = \bar{X} = \frac{1}{N}\Sigma_{i=0}^{N} x_i$, that is the sample mean (sample average), the standard normal for this RV can be written [67]:

$$Z = \frac{\hat{\mu} - \mu}{\frac{\sigma}{\sqrt{n}}} \tag{2.22}$$

It is immediately obvious that there is something weird going on in the denominator and the since theoretically both μ and σ are unknown, it is not trivial to calculate Z. The denominator results from the fact that a variance of an estimator that is normally distributed is different from a variance of a regular normal random variable, i.e. $\mathbb{V}[X] \neq \mathbb{V}[\bar{X}]$ and $\mathbb{V}[\bar{X}] = \frac{\sigma}{\sqrt{n}}$ where n is the sample count. This immediately indicates that the variance of the estimator is decreasing with increasing sample size, i.e. the more samples one gets the more certain given statement is. With that, one can reach one of the most crucial inventions of the frequentist statistics - sampling distributions. They are, basically, distributions for a given calculated statistic (again: sample median, sample mean, sample standard deviation etc.) that describe how they would vary upon resampling. Thus a sampling distribution for a sample mean can be written as $\bar{X} \sim N(\hat{\mu}, \frac{\hat{\sigma}}{\sqrt{(n)}})$, which results in the form of the Standard Normal calculated from it.

Having introduced sampling distributions, it would be immensely practical to be able to use them to make statements with a degree of certainty about the possible values of the parameter. It is, after all, one of the primary goals of inferential statistics. To this end, Confidence Intervals are utilized. To begin to construct such interval one starts with a condition [91]:

$$P(-1.96 \leq Z \leq 1.96) = 0.95 \tag{2.23}$$

The statement simply reads that between the values -1.96 and 1.96, for a Standard Normal distribution, 95% of probability is accumulated. Upon substitution of the definition of the Standard Normal for the sample mean and rearragnement [91]:

$$P(\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{(n)}} \leq \mu \leq \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{(n)}}) = 0.95 \tag{2.24}$$

This is the definition of a 95% Frequentist Confidence Interval (FCI) on the population mean [91]. This shows that by evaluating the insides of the probability operator, one can obtain the range within which there is 95% of probability placed on the mean. This requires careful interpretation: in frequentist approach this does not mean, that there is 95% certainty that the population mean is located within the specified range. This specification means that upon resampling, i.e. completing multiple experiments, 95% of the time the population mean would be located inside the FCI. It can be calculated for other confidence values, α_{FCI}, by replacing the ± 1.96 in the formula, by a quantile from the Standard Normal distribution corresponding to a given confidence level, $z_{\frac{\alpha}{2}}$, when a two-sided FCI is sought after. [91] The FCI can be calculated for other parameters than the mean and for parameters that represent other distributions (e.g. Bernoulli). The definition in Eq. 2.24 makes two crucial assumptions that are rarely met in reality: the sample size is large enough to use the Standard Normal distribution quantiles and σ is assumed to be known.

Not knowing σ is not an obstacle when n is very large, then Z approximates the true Standard Normal very well by assuming $\sigma = S = \hat{\sigma}$, where S is the sample standard deviation. Trouble starts when n drops, the approximation quickly fails to be true. [67] To solve this issue a T random variable (as opposed to Z)

is introduced and defined as [67]:

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{(n)}}} \tag{2.25}$$

As opposed to Z being normally distributed, $Z \sim N(0,1)$, T (called often the t-statistic) is t-Student distributed, $T \sim t-Student(\mu, S, \nu)$, where $\nu = n-1$ is called the 'degrees of freedom' often termed the 'normality parameter' [67, 92], as when $\nu \to \infty$, $t - Student \to Normal$. Now, repeating the same operations as above and allowing for an arbitrary confidence level, α:

$$\mu \in \bar{X} \pm t_{\frac{\alpha}{2}, n-1} \cdot \frac{s}{\sqrt{n}} \tag{2.26}$$

2.4.2 Null Hypothesis Significance Testing

A statistical hypothesis is a statement made about a value of a (most often) single parameter. [67]. Any hypothesis needs to be stated, investigated and then - based on data - accepted or rejected. In Null Hypothesis Significance Testing (NHST), there are usually two hypotheses that one cares about: H_0 and H_A. The first one, H_0, called the null hypothesis is the base hypothesis upon which the current state-of-the-art knowledge is based. H_0 is often established through years of research in the classical sense, in other cases it may be a personal choice about some assumption or underlying truth (e.g. in Regression Analysis, on that later, see Section 2.4.3). H_A is the alternative hypothesis, the one which is compared against H_0, and the one which leads to either rejection or failure to reject H_0 (using the correct statistical parlance). A example way to pose a NHST problem mathematically may be seen as [91, 92]:

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_A &: \theta \neq \theta_0 \end{aligned} \tag{2.27}$$

which can be interpreted as H_0 states that some parameter θ (e.g. the population mean) is equal to some value, θ_0 and the alternative hypothesis, H_A disagrees. There are two other commonly used settings, i.e. $H_A : \theta > \theta_0$ or $H_A : \theta < \theta_0$. Having defined the types of hypotheses and statements made

about them, it is important to now show a procedure to either proving or disproving them. Going straight to a more general case, when sample count is not necessarily large and one can't rely on the Standard Normal approximation, the t-Statistic from equation 2.25 is reused. Furthermore two errors are defined:

- Type I Error: H_0 is rejected despite being true
- Type II Error: H_0 is not rejected despite being false

NHST deals with the first error type. To evaluate the veracity of statements, the following logic is assumed [91]:

$$P(\text{type I error}) = P(T \geq t_{alpha,n-1} \text{ when } T \sim t - Student(\nu)) = 1 - \alpha \quad (2.28)$$

P(type I error) is often called α. It is chosen by the experimenter before conducting an experiment to prevent any foul play. For a given selected α, probabilites resulting from the above equation are termed p-values. To calculate a p-value for the two-tailed test (the current setting) first has to calculate the t-Statistic from Equation 2.25. Then to calculate the p-value:

$$p - value = 2 \cdot (1 - F_{t_\nu}^{-1}(T)) \quad (2.29)$$

where $F_{t_{nu}}^{-1}(T)$ is the inverse Cumulative Distribution Function (CDF) of the t-Student's distribution at the t-Statistic value. When the calculated p-value is lower than the assumed threshold (α, most often chosen to be 0.05) the null hypothesis is rejected, otherwise a failure of rejection is reported. It is important to note that when testing multiple hypotheses, the problem of multiple comparisons arises and needs to be accounted for. [93]

2.4.3 Statistical inderence for Linear Regression

The goal of Standard Linear Regression (SLR, also termed Ordinary Least Squares - OLS, here used interchangeably) is to be able to inter- and extrapolate some response, Y, based on a chosen set of predictors, x, i.e. $Y = f(x)$. From a statistical perspective there is one additional thing - the error term, ϵ, resulting in $Y = f(x) + \epsilon$. The error term itself is normally distributed with a mean equal to zero and constant variance, σ^2, i.e. $\epsilon \sim N(0, \sigma^2)$. Such a construction of Y induces a normal distribution :

$$Y \sim N(\mu, \sigma^2), (i.id.) \tag{2.30}$$

where μ is the mean response and σ^2 is the variance. One crucial detail is that the data, Y are assumed to be "iid" (alternatively denoted i.id.) which stands for 'independent and identically distributed'. This means, that each datapoint is generated independently of another and that they follow a similar distribution. It is a common and often implied assumption in statistical modeling, which applies also in context of statistical intervals and hypothesis testing mentioned above. Going further,the mean response can be rewritten in terms of x as:

$$\mu = \mathbb{E}[Y] = E[f(x) + \epsilon] = f(x) \tag{2.31}$$

which is due to the fact that x's are not treated as random variables and the mean of ϵ is 0. Subsequently also the variance is:

$$\sigma^2 = \mathbb{V}[Y] = \mathbb{V}[f(x) + \epsilon] = \mathbb{V}[\epsilon] = \sigma^2 \tag{2.32}$$

again due to the same reasons as above. In the context of Ordinary Least Squares (OLS), f(x) is often rewritten in terms of a linear combination of the predictors, x, and some paramerters, β, which are then found given x and Y: [91]

$$f(x) = \beta_0 + \Sigma_{i=1}^{p} \beta_i x_i \tag{2.33}$$

where p is the amount of predictors and β_0 is the intercept term. The interpretation of the parameters (in Multiple Linear Regression problems) is as follows: [94]

- β_0 average of all Y when all x are at their 0
- β_j average change in Y associated with a unit increase in the value of x assuming all other predictors are held constant (controlled for)

The parameters may be obtained through many ways: Maximum Likelihood Estimation, usage of normal equations or other optimization procedures [8]. In statistical inference for linear regression, the model parameters themselves are

assumed to be random variables resulting in them having their own sampling distribution: [94]

$$\hat{\beta} \sim N(\beta, \sigma^2(X^T X)^{-1}) \tag{2.34}$$

First, one may perform hypothesis testing on βs to estimate whether they are statistically significant using NHST. The hypothesis is posed as:

$$\begin{aligned} H_0 &: \beta = 0 \\ H_A &: \beta \neq 0 \end{aligned} \tag{2.35}$$

the t-Test statistic is defined as:

$$T_i = \frac{\hat{\beta}_j}{\sqrt{[\hat{\sigma}^2(X^T X)^{-1}]_{jj}}} \tag{2.36}$$

where $\hat{\sigma}^2$ is the estimator for the noise in the model and is defined as $\hat{\sigma}^2 = \frac{RSS}{n-(p+1)}$. Using the t-statistic we may perform NHST to be able to tell which model parameters are statistically significant (i.e. p-value $< \alpha$) and which are not. When many parameters are present in the model it is relatively easy to inflate the type I error rate, as mentioned in previous sections, leading to insignificant parameters being deemed significant. To solve this issue various approaches have been developed like conducting an F-test [91] or using Bootstrap Inclusion Frequencies [95]. The former compares the full model's explanatory power to the one with only an intercept term, while the latter involves using resampling with replacement (bootstrapping) of the data and refitting the model and subsequently tallying up the amount of times a given parameter was deemed significant.

It is also possible to develop CIs for βs to be able to assess their variation. A $(1-\alpha)\cdot 100\%$ Confidence Interval is defined as: [94]

$$\beta_{true} \in \hat{\beta}_j \pm t_{\frac{\alpha}{2}} \cdot (n-(p+1))\sqrt{[\hat{\sigma}^2(X^T X)^{-1}]} \tag{2.37}$$

In statistical analyses it is also common to interpret and report Prediction Intervals, PIs. As opposed to Confidence Intervals which are calculated for

estimated parameters, PIs are calculated for actual, often, future data. A $(1-\alpha)\cdot 100\%$ Prediction Interval for an unobserved predictor vector x^* for an OLS estimator is defined as: [94]

$$y_i^* \pm t_{\frac{\alpha}{2}} \cdot (n-(p+1))\sqrt{[\hat{\sigma}^2 x^*(X^T X)^{-1}](x^*)^T} \tag{2.38}$$

where y_i^* is some unobserved data point. The Prediction Intervals are always wider or equal to Confidence Intervals.

2.4.4 Assumptions for valid inferences in SLR

To be able to make statistical statements about the prediction (its quality, parameters etc.) it is crucial that a given SLR model meets the following assumptions [67]:

- **Linearity**: model must be linear in parameters, β
- **Independence**: error terms, ϵ are independent of each other
- **Homoscedasticity**: error terms, ϵ, have constant variance σ^2
- **Normality**: errors are normally distributed, $\epsilon \sim N(0, \sigma^2)$

When any of the mentioned assumptions is not met, the fitted model may result in a biased estimator for β which leads to poor explanatory power and wrong inferences about both the model and the data. Moreover it is also possible that the model's predictions will be less accurate and misleading. The most immediate and simple way to visualize these assumptions is to plot them, especially to plot the examples of glaring violations of these principles as done in Figure 2.5. The methods present in the plots (along with some others recommended by literature [67]) are listed and described below:

- **Linearity**: Fitted vs Observed Data plot - assesses the fit to the data and informs whether the relation observed data is actually linear (see Figure 2.5, **A**)
- **Independence**:
  - Residual vs Time Index plot - shows whether there is some unresolved structure in the residuals which potentially could mean that they are dependent (see Figure 2.5, **B**)

  - Successive Residual Plot - plots residuals against each other if there is some trend present it means that the residuals are in fact correlated
  - Durbin-Watson test - assesses successive correlations between residuals

- **Homoscedasticity**: Residual vs Fitted Value Plot - if there is an increase in dispersion of residuals for successive values it means that the variance is not constant (see Figure 2.5, **C**)
- **Normality**:
  - Quantile-Quantile plots - compare the quantiles of the residuals versus the quantiles of a similar normal distribution, if large discrepancies are present it means that the normality assumption is violated (see Figure 2.5, **D**)
  - Shapiro-Wilk test for normality tests whether the given data do actually come from a normal distribution

2.4.5 Fit of statistical models

Having met all conditions for statistical inference, it is also very important that the model is realistically a good representation of the data under consideration. Many robust criteria have been developed assess the 'goodness of fit' of the obtained model. Akaike Information Criterion (AIC) and Baysian Information Criterion (BIC) are one of the most popular ones as they also account for the parameter count, which when very large, artificially inflates the fit quality. AIC is defined as [96]:

$$AIC(\eta(X|\hat{\beta})) = 2 \cdot (p+1) - 2logL(\hat{\beta}) \tag{2.39}$$

where $\eta(X|\hat{\beta})$ is the fitted model, p is the parameter count and $L(..)$ is the likelihood function. The right hand side is a goodness of fit measure while the left hand side is the complexity penalty. Similar to AIC is BIC, defined as [96]:

$$BIC(\eta(X|\hat{\beta})) = (p+1)log(n) - 2logL(\hat{\beta}) \tag{2.40}$$

it is obvious that it is analogous to AIC, however it differs in a penalty, here it is much stricter. Generally AIC and BIC do not have an 'ideal' set of values

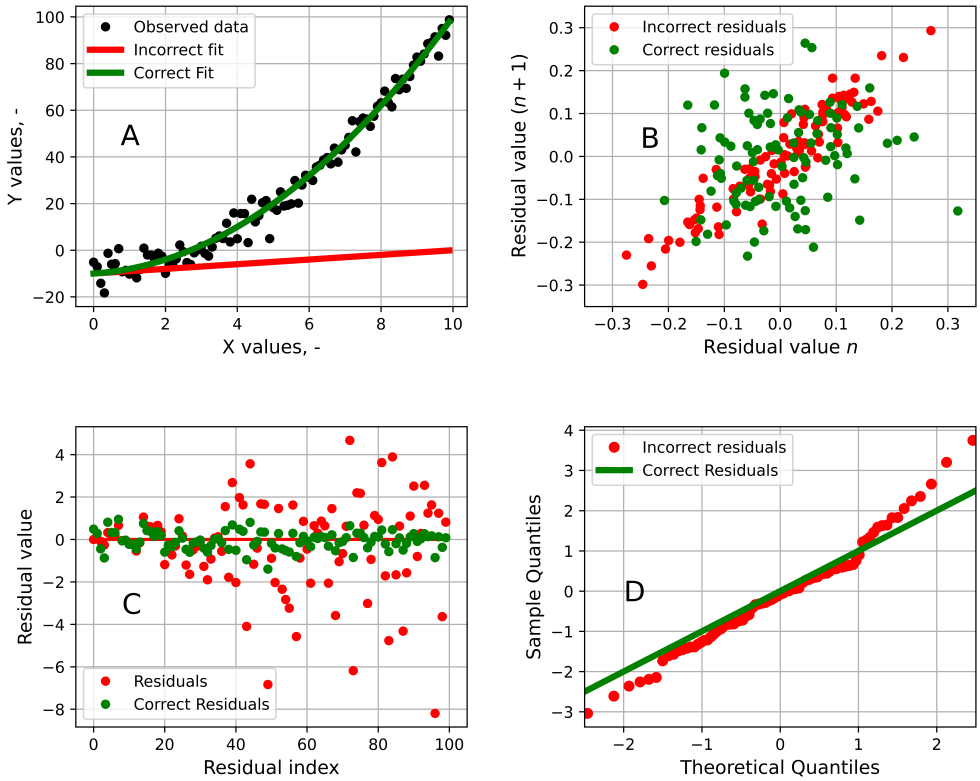

Figure 2.5: Exemplary plot showing violations and fulfillment of the four fundamental assumptions in classic OLS (red) along with what an ideal behaviour of OLS would look like (green), generated using artificial data. **A**: Violation of the linearity assumption. **B**: Violation of the independence assumption. **C**: Violation of homoscedasticity assumption. **D**: Violation of the normality principle.

for a model, it's best when they are as small as possible and while comparing multiple models, the one with the smallest AIC and BIC is possibly the best one, meaning that it has the best trade-off between parameter count and explanatory power (however it is not guaranteed that AIC and BIC will choose the same models). One of the most robust tehcniques for model comparison is the Likelihood-Ratio test (LRT). It is defined as: [91]

$$\lambda_{LR} = -2[logL(\theta_0) - logL(\hat{\theta})] \tag{2.41}$$

where θ_0 and θ represent some parameters of two different models (hypotheses). $\lambda_L R$ is asymptotically χ^2 distributed if the Null is true. LRT is small when the alternative model is better than the Null one. Moreover it is guaranteed to be the most powerful test among all tests of α significance level by the Neyman-Pearson Lemma. [97]

Aside for assessing model fit, it is important to keep in mind it's performance on new, unseen test data. Starting off with Root-mean Squared Error (RMSE), the definition is: [98]

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.42}$$

where y represents the column vector of observed data and $\hat{y}$ represents the column vector of predictions, with n being the number of samples. RMSE is particularly sensitive to outliers because it involves squaring the differences and then averaging them. The Mean Absolute Error (MAE) is computed as follows: [98]

$$MAE(y, \hat{y}) = \frac{1}{n}\Sigma_{i=1}^{n}|y_i - \hat{y}_i| \tag{2.43}$$

where $|..|$ represents the absolute value. MAE is generally more robust against outliers. Median Absolute Error, also known as MEDS or MedAE, is described as: [99]

$$MEDS(y, \hat{y}) = median(|y_1 - \hat{y}_1|, |y_2 - \hat{y}_2|, ..., |y_n - \hat{y}_n|) \tag{2.44}$$

The robustness of MEDS against outliers is enhanced by its use of the absolute metric and the employment of the median rather than the mean of the errors calculated. The Maximum Absolute Error (MaxErr) is characterized in: [100]

$$MaxErr(y, \hat{y}) = max(|y_1 - \hat{y}_1|, |y_2 - \hat{y}_2|, ..., |y_n - \hat{y}_n|) \tag{2.45}$$

This signifies the most unfavorable outcome in terms of the efficacy of the model. The Mean Relative Error for RMSE is characterized as:

$$MRE = \frac{RMSE(y, \hat{y})}{\hat{\mu}_y} \cdot 100\% \tag{2.46}$$

where $\hat{\mu}_y$ represents the average of the variable in question. The RMSE metric can be substituted with any suitable metric. Relative errors are given to offer a scale of the error values relative to the values found in the test set.

2.5 ANOVA in linear regression perspective

2.5.1 One-Way ANOVA and the F-test

ANOVA stands for ANalysis Of VAriance. It's primary goal is to find an answer to the question "Are there differences with respect to the mean of a (continuous) variable across treatments or groups?". In simple terms, the goal is to examine whether there are some statistically significant differences between groups in a given experiment [101]. A One-Way ANOVA takes a look at the difference between a given factor across groups. Formally it can be written as an ANOVA model [101]:

$$y_{ij} = \mu_j + \epsilon_{ij} \tag{2.47}$$

where y_{ij} is some i-th response of j-th group, μ_j is that group's mean and ϵ is the noise. Here the assumption is that a given $y_{ij} \sim N(\mu_j, \sigma^2)$, i.e. individual groups come from different distributions but all share the same variance [101]. This can be restated as an effects model:

$$y_{ij} = \mu + \alpha_j + \epsilon_{ij} \tag{2.48}$$

here μ is sometimes called the 'grand' or 'population' mean and is shared across all groups, while α_j is the departure of group j from the overall mean, so called **effect**. Both of these formulations assume that errors share the same, constant variance across groups, i.e. $\epsilon \sim N(0, \sigma^2)$. The decomposition of variance in ANOVA occurs by defining three measures: Total Sum of Squares (TSS), Treatment Sum of Squares (SSTr) and the Residual Sum of Squares (SSR). The TSS is defined as:

$$TSS = \Sigma_{j=1}^{J} \Sigma_{i=1}^{n_i} (Y_{ij} - \bar{y}_{..})^2 \tag{2.49}$$

where J is the group count, n_i is the number of observations in each group and $\bar{y}_{..}$ is the Grand Mean, i.e.: the total mean over all samples and groups. It is defined as:

$$\bar{y}_{..} = \frac{1}{n_j}\frac{1}{n_j}\Sigma\Sigma_{i=1}^{n_j} Y_{ij} \tag{2.50}$$

It is also useful to introduce the group mean as:

$$\bar{y}_{.j} = \frac{1}{n_i}\Sigma_{i=1}^{n_i} Y_{ij} \tag{2.51}$$

Having a look at SSTr:

$$SSTr = \Sigma_{j=1}^{J}(\bar{y_{.j}} - \bar{y}_{..})^2 \tag{2.52}$$

one can see that it is the sum of squared deviations of the group level means from the overall, grand mean. Finally SSR is defined as:

$$SSR = \Sigma_{j=1}^{n_j}(Y_{ij} - \hat{y}_{.j})^2 \tag{2.53}$$

Furthermore it can be shown ([101]) that:

$$TSS = SSTr + SSR \tag{2.54}$$

i.e. the Total Sum Of Squares can be decomposed into the sum of individual-level variations and group-level variations. Similar to OLS an F-test is applied, which is a common test that allows to compare different models based on how much of the variance they explain. [67] Unlike in OLS, itss definition of the F-statistic is a little different:

$$F_{ANOVA} = \frac{\frac{SSTr}{J-1}}{\frac{SSR}{n-J}} \tag{2.55}$$

this induces a an F distribution on the statistic, $F_{ANOVA} \sim F_{(J-1),(n-J)}$. The hypothesis tested here is as follows:

$$\begin{aligned} H_0 &: \mu_j = \mu_k, j \neq k \\ H_A &: \mu_j \neq \mu_k, j \neq k \end{aligned} \tag{2.56}$$

i.e the Null Hypothesis states that there are no significant differences between given groups, while H_1 claims the converse. What follows is an F-test and upon rejection of the Null, it is possible to further examine the data for pairwise differences and other desired information. [67].

2.5.2 Two-way ANOVA

To further extend ANOVA to be able to compare data where more than one factor are considered, Two-Way ANOVA was created. Again it can be stated as the 'means' and 'effects' models, where the former is defined as [67, 101]:

$$Y_{ijk} = \mu_{jk} + \epsilon_{ijk} \tag{2.57}$$

where μ_{jk} are the means of intersecting groups and $\epsilon_{ijk} \sim N(0, \sigma^2)$ is the error term. The 'effects' notation is much more transparent:

$$Y_{ijk} = \mu + \tau_j + \alpha_k + \epsilon_{ijk} \tag{2.58}$$

where both τ and α are effects. τ_j is interpreted as the deviation of the response in the j-th level of the τ factor, while holding the α factor at its mean value. [101] Due to the fact that this notation explicitly states the effects it is much more common in the statistical literature, particularly concerning multilevel (random/mixed effects) models [102]. Such explicit notation also makes it easier to model possible interactions between different effects and impose a hierarchical structure on data [102].

When designing an actual scientific study where the goal is to discriminate against factors effects between groups (eg. A/B testing) it needs to be decided how the factors relate to each other. There are many design approaches, but the two particularly relevant for this work are crossed and nested designs. A fully crossed design occurs when all factors are jointly represented at all levels (i.e. all factor-level combinations are measured). This design can be presented as a table, see 2.1. From the Table 2.1 (a) it is obvious to see that there are all 6 possible measured combinations (indicated by ones) of the factor's levels $((\alpha_1, \tau_1), (\alpha_1, \tau_2), (\alpha_2, \tau_2)$ and so on). A different kind of design, is one where factors are nested within each other. Such kind of design is simply called nested and it can be also presented as a table, see Table 2.1 (b). Here the τ_1 level of factor τ is nested inside α's first level. This means that not all factor-level

Table 2.1: Two fundamental design philosophies for two-way ANOVA for two factors τ and α. **a**: crossed effects design, i.e. each factor-level combination is present in the design. **b**: Nested design: some factors are nested within others, preventing all factor-level combinations to be present.

Factor Levels	α_1	α_2	α_3
τ_1	1	1	1
τ_2	1	1	1
τ_3	1	1	1

(a) Crossed design

Factor Levels	α_1	α_2
τ_1	1	0
τ_2	1	0
τ_3	0	1
τ_4	0	1

(b) Nested design

combinations are present (i.e. there is no combination for the second level of α with the τ's first level).

Although it may seem difficult to picture, both designs can be readily understandable. Crossed designs occur when all subjects are tested with all possible factors. A relevant example, somewhat similar to [103], could be testing human subjects for impact of exercise on the measured arterial stiffness (through PWV) where α could be distributed over levels corresponding to age brackets (e.g. young, adult, middle-aged etc.) and τ could be the exercise type (e.g. running, cycling etc.), where each person in each age bracket took part in each of the exercises, thus exhausting all the factor-level combinations. Nested designs occur when one factor is nested inside another (or some of factors levels are nested in some). An example can be based on [104], where the response to a given treatment is being sought and the nested factor (τ) is the treatment itself and the nesting factor (α) is the country of origin or region of origin. As a person cannot usually come from two countries at a time or at least be measured in two different countries, their respective response to treatment is forever nested in a country of origin factor. Nested designs can occur artificially or naturally. Natural occurrence is when the physical processes that causes the scientist to conduct experiments are implicitly nested (e.g. country of origin, students nested within schools nested within regions [102]), while artificial designs arise when the nesting structure is constructed by the scientist during experimental design.

2.5.3 ANOVA recast as Linear Regression

Going straight to the two-way ANOVA it is possible to recast it as a linear regression problem. The purpose for such an operation is obvious: having developed a large body of techniques and assumptions for SLR it would be beneficial to extend them to ANOVA to facilitate its application. To do this, consider two factors, α and τ. Let's say α is distributed over three levels and τ is distributed over 2 levels, as before. The linear regression model for such a case would be written as follows:

$$y_i = \beta_0 + \beta_1 \tau_{2,i} + \beta_2 \tau_{3,i} + \beta_3 \alpha_{2,i} + \epsilon_i \tag{2.59}$$

where the notation $\tau_{j,i}$ means the i-th observation of the j-th factor level. The first thing to notice here is the absence of $\alpha_{1,i}$ and $\tau_{1,j}$. It is not accidental - for stability and well-posedness issues, both of these factor levels are absorbed into the β_0 coefficient serving as a baseline. The interpretation of β_0 thus changes from the one in OLS (see section 2.4.3) to be the average value of the response Y at the first level of each of the factors. For this simple case, a full interpretation of regression coefficients and how they relate to factors is presented in Table 2.2.

Table 2.2: The interpretation of linear regression coefficients in context of effects in a two-way ANOVA regression framing.

Factor Levels	α_1	α_2	α_3
τ_1	β_0	$\beta_0 + \beta_1$	$\beta_0 + \beta_2$
τ_2	$\beta_0 + \beta_3$	$\beta_0 + \beta_1 + \beta_3$	$\beta_0 + \beta_2 + \beta_3$

Here it can be seen that the baseline scenario is the level one of both τ and α. In experimental setting it would often be the control group (i.e. no applied treatment or a placebo pill etc.), whereas each subsequent entry in the Table is the deviation from the baseline group. By framing ANOVA from the OLS viewpoint, all the tests and methods that were discussed previously (see Section 2.4.3) now are also easily applied.

2.6 Linear Mixed-Effects Modeling

Mixed-effects models are a statistical approach that combines random and fixed effects. They are most often used to model hierarchical (or nested) data. Such a hierarchical structure may arise in various ways e.g. through performing multiple measurements on a repeated setup (or subject) but also from spatial and/or temporal structure. [105] At the lowest rung of the hierarchy there are individual observations, this is called the data or unit level [105]. Each subsequent level above the lowest one may be modeled by fixed and/or random factors. For models defined based on the effects approach, 'fixed and random factors' become 'fixed and random effects' and that is the nomenclature that will be followed for the rest of the work. Fixed effects have been used thus far: they are regular factors that are treated at face value and no variation for them is assumed. Random effects, on the other hand, assume that the currently considered set of predictors is just a sample from some distribution and therefore are prone to exhibit variation. The great thing about them is that, while using fixed effects one is constrained to inferences regarding the measured levels, for random effects it is possible to extend the inferences towards all possible, unmeasured levels. [106] More formally random effects are effects that are estimated at each factor level, but where the distribution of these estimates is explicitly modeled by hyperparameters [105], i.e. parameters of the estimated distribution from which they come. It can be easily visualized by having a look at the simplest possible random-effects model, the random intercept model:

$$y_{ij} = \beta_{0j} + \epsilon_{ij} \tag{2.60}$$

where j is the group index and i the individual unit level index and $\epsilon_{ij} \sim N(0, \sigma_R^2)$. Here β_{0j} is the intercept, however it varies from group to group. This means that each group has its own estimated β_{0j}, but there is also an overall mean. β_{0j} may be deconstructed as [102]:

$$\beta_{0j} = \beta_0 + \rho_{0j} + \epsilon_{ij} \tag{2.61}$$

where β_0 is the overall group-level mean nad ρ_{0j} is the individuals groups deviation from it, $\rho_{0j} \sim N(0, \sigma_j^2)$. Thus the group-level mean comes from a distribution $\beta_{0j} \sim N(\beta_0, \sigma_j^2)$ [107] and its variance may be considered as explained in the sense of the model (i.e. it takes away from the residual variance σ_R^2), but in the sense of a given level it is unexplained. [105] There are many possible

basic mixed-effects models and their application and usefulness can be seen by looking at Figure 2.6. Figure 2.6 A shows a cloud of data and immediately one advantage of the grouping approach - having a look at it, no obvious trend can be seen when the data is pooled and just an intercept model is fitted. However having a look at Figure 2.6 B, individual groups indicate some patterns in the data and by using a very basic mixed effects model (namely random intercept, fixed slope model) one can see that the dependencies are somewhat resolved. The formula present on the plot, i.e.:

$$y_{ij} = \beta_{0j} + \beta_1 X_{ij} + \epsilon_{ij} \tag{2.62}$$

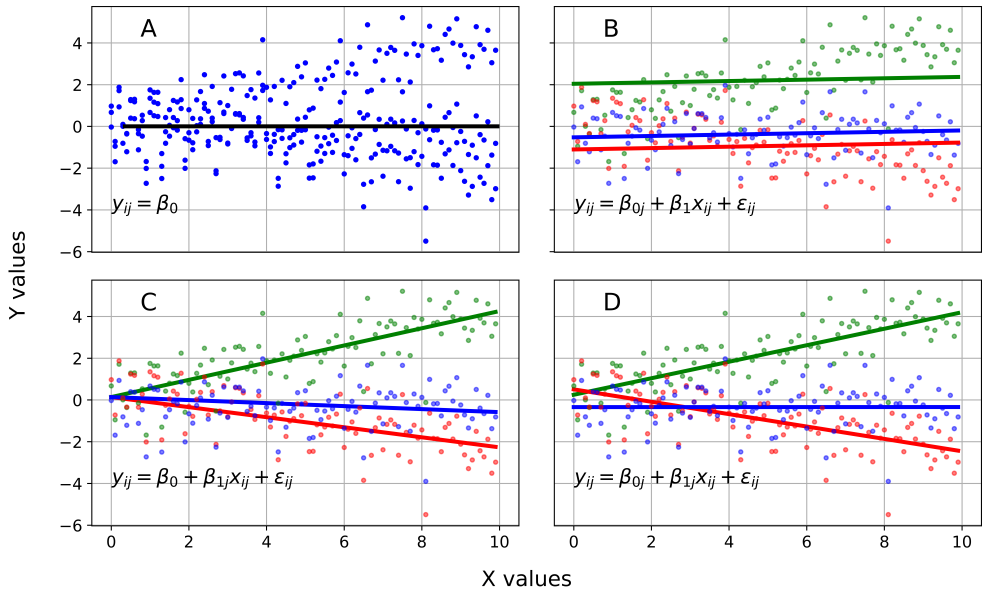

Figure 2.6: Example of various possible mixed effects models for a single data cloud. **A**: a regular fixed-effects with an intercept misses any possible trend in data. **B**: By enable a random intercept it is already possible to see some group differences. **C**: Varying intercept model captures the trend well but not the average value. **D**: a fully random-effects model manages to capture both group averages and trends present in the data. Generated using artificial data.

models the intercept as a random effect coming from a distribution and β_1, i.e. the fixed effect, without estimating any hyperparameters related to it. Even

though this approach is better, it is not the answer to the problem - while it captures the differences in group means, it does not capture its trend. The converse statement can be made about the varying slope model, i.e:

$$y_{ij} = \beta_0 + \beta_{j1} X_{ij} + \epsilon_{ij} \tag{2.63}$$

where this time the slope is a random effect and the intercept is fixed. Here the average change in y with respect to X is captured very well, but not the average group value. This problem is solved by considering a full random effects model. This can be seen in the bottom right of Figure 2.6. It is a model with both a varying intercept and a varying slope:

$$y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \epsilon_{ij} \tag{2.64}$$

For clarity, it is possible to completely unravel the fully random-effects model by describing the coefficients:

$$\begin{aligned} \beta_{0j} &= \beta_0 + \gamma_{0j} \\ \beta_{1j} &= \beta_1 + \gamma_{1j} \end{aligned} \tag{2.65}$$

where β_0 and β_1 are the fixed effects and γ_{0j} and γ_{1j} are the random effects. All of the models with random-effects present were effectively two-level models, where the lower level was the unit level and the higher level was the coefficients' ('group') level. It is possible to extend these models to many more levels and to even model coefficients as themselves dependent on some measured predictor Z (in an analogous manner to how Y is dependent on X). [102]

2.6.1 Benefits of using mixed-effects and multilevel models

As seen in Figure 2.6 the primary benefit of using linear mixed-effects models (LMEs) is their ability to extract relationships and structure from data, that would otherwise be missed had a pooled model been assumed. [103] Another reason to use LMEs is when correlated errors are present. The assumption is that the unit in group j would be more similar to another unit in that group, than to a unit in a different group, i.e. more correlated. This breaks the independence assumption present in OLS and needs to be accounted for until

one can again assume $\epsilon_{ji} \sim^{iid} N(0, \sigma^2 R)$. Omitting these effects could potentially underestimate the standard error on the mean coefficient, inflate the test statistic and lead to decreased p-values and committing a Type I error far more often than the assumed rate. [102] Having a look at hierarchical (multilevel) models, the primary benefit for using them is the phenomenon called 'variance shrinkage'. It causes the group-level estimates to 'shrink' towards the population mean. It happens because the group-level estimates are now no longer only influenced by their individual groups, but also by the population mean. [105, 102] This is particularly pronounced for models where certain groups have a very small amount of observations. In such cases the groups estimated are 'pulled' towards higher level estimates, even when the sample count is dramatically low (e.g. one or two datapoints). [105, 92] It is important to keep in mind, that LMEs are not a panaceum for each dataset and they definitely need to be carefully studied before applying.[103, 77, 102] Nevertheless, many scholars strongly encourage to at least consider them before moving on to either traditional OLS or some other techniques (like Generalized Estimating Equations [108]).

2.7 FSI Governing Equations

The governing equations for the Fluid-Structure interaction problem are customarily derived from the Cauchy equation of motion. The dynamic mechanical behavior of the fluid domain is governed by the Navier-Stokes equations, i.e., the conservation of momentum and mass equations. The conservation of momentum can be expressed as follows: [109]

$$\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla)\vec{v} = -\frac{1}{\rho}\nabla p + \frac{\mu}{\rho}(\nabla^2 \vec{v}) + \frac{1}{\rho}(\frac{\mu}{3} + \kappa)\nabla(\nabla \cdot \vec{v}) + \vec{b} \tag{2.66}$$

where $\vec{b}$ are the body forces acting on a fluid (N), ρ is the density (kg/m3), $\vec{v}$ is the velocity vector field (m/s), μ and κ are the first and second Lame constants and ∇ is the nabla operator, i.e. $\nabla = \Sigma_{i=0}^{N} \frac{\partial}{\partial x_i}\vec{e}_i$ ($\vec{e}_i$ being a general unit vector). For incompressible fluids, the conservation of mass equation can be written as [110]

$$\nabla \cdot \vec{v} = 0 \tag{2.67}$$

The solid domain is also modeled using Cauchy's equation of motion: [9]

$$\rho_s \frac{\partial \vec{U_s}}{\partial t} - \nabla \cdot \overleftrightarrow{\sigma}_s = \vec{b}_s \tag{2.68a}$$

Where $\vec{U_s}$ is the displacement, $\overleftrightarrow{\sigma}_s$ is the Cauchy Stress Tensor. For linearly elastic bodies it is assumed that the tensor is solely dependent on the Green strain tensor, which under small deformations, simplifies to.

$$\sigma_{s,ij} = 2\mu\epsilon_{ij} + \lambda\epsilon_{kk}\delta_{ij} \tag{2.69}$$

where ϵ_{ij} and $\sigma_{s,ij}$ indicate the components of the strain and Cauchy stress tensors, respectively, μ is the shear modulus (or viscosity), λ is Lame's constant. δ_{ij} is the Kronecker delta and index repetition implies summation. The base material used in ANSYS simulations is the linear elastic material with large deflection turned on. It is the default and recommended material to use in ANSYS documentations as it allows for actually displacing the mesh of the solid body, thus in turn allowing to displace the fluid domain and complete the Two-Way coupling. [13] In contrast to linearly elastic materials, in hyperelastic materials, the imposed strain on the bodies can be stored in the form of energy (the so-called 'deformation potential work'). One of the most commonly used material definitions in hyperelasticity is the Neo-Hookean material, with its strain energy density function defined as [109]:

$$\Phi = \frac{\mu}{2}(I_1 - 3) - \mu \ln J + \frac{\lambda}{2}(\ln J)^2, \tag{2.70}$$

where $I_1 = \text{tr}\ (\mathbf{C})$ and $J^2 = |\mathbf{C}|$ and C is the right Cauchy Deformation Tensor. tr(...) stands for the trace operation which is summing the elements on the matrix main diagonal, i.e. for some square matrix A, $A \in \mathbb{R}^{kxk}$: $tr(A) = \Sigma_{i=0}^{k} A_{ii}$ From it an expression for the Cauchy's Stress Tensor is often described in terms of the left Cauchy Deformation tensor, B, and may be derived from the Second Piola-Kirchoff tensor [111] to be:

$$\overleftrightarrow{\sigma_s} = \frac{\mu}{J}(\overleftrightarrow{B} - \overleftrightarrow{I}) + \frac{\lambda}{J}(\ln(J))\ \overleftrightarrow{I} \tag{2.71}$$

where $\overleftrightarrow{I}_s$ is the identity tensor. The discretization procedure of the equations for coupled equations is described in detail e.g. in [12] Another commonly used model is the Isotropic Elastic (IE) model, defined as [112]:

$$\phi = \frac{1}{2}\lambda(tr\ \overleftrightarrow{E})^2 + \mu\ \overleftrightarrow{E}:\overleftrightarrow{E} \tag{2.72}$$

where $\overleftrightarrow{E}$ is the Euler-Lagrange tensor. This material reduces to a regular linear elastic model when small deformations are present, but is applicable to large deformations and rotations. [12]

3

Surrogates, UQSA and Bayesian Inference

This Chapter covers the theory of approaches and techniques applied in Chapter 5. It begins by describing the background behind experimental design and its modern variant - sampling schemes. Then it moves on to describe one of the two surrogates developed in the thesis - Gaussian Process Regression. It describes in depth its roots, modeling approach and the way to perform inference. It also tries to convey the intuition and logic behind GPR's implementation of kernel functions (kernels). It then moves on to describe how the kernel hyperparameters are optimized and finishes with a brief description of GPR variants. Then the second of the developed models is described - Singular Value Decomposition-based (SVD) Reduced Order Model (ROM). The theory behind SVD is introduced to help make sense of the data order reduction procedure. Further, one of the applications of the surrogates is described - Uncertainty Quantification and Sensitivity Analysis. The variance-based methods are described in detail along with First Order and Total Sobol Indices and their Monte Carlo estimates. Last, but not least, is the introduction of Bayesian Statistics. There is a brief reminder of the Bayes Theorem. What follows is the description of metrics used in Bayesian Statistics for describing uncertainty as well as methods used for fitting Bayesian models. Finally, the diagnostics for quality checking a Bayesian model are introduced.

3.1 Design of physical and virtual experiments

In experimental design, the key thing is to learn as much as possible about the system the researcher is trying to understand with the fewest possible number of experiments run. A given experiment is often characterized in terms of factors, i.e. variables one is interested in (e.g. pressure, temperature) and levels, i.e. values these factors can assume (e.g. discrete "low and high" values or a given discrete range). When one is interested in observing an effect, i.e. the response of a system due to a stepwise change in a factor's value [94], one needs to perform an experiment when a given factor is varied, while all the others stay fixed. This naturally leads to an explosion of measurements that need to be carrried out, as for a very simple case of K factors distributed over 2 levels, i.e. assuming only two values, the number of experimental evaluations, $N_{measurements}$, necessary to individuate all of the effects grows as:

$$N_{\text{measurements}} = 2^K \tag{3.1}$$

(and more generally as M^K for K factors distributed over M levels). This is termed the Full Factorial Design [94]. Sadly, when it comes to physical experiments, various limitations to the amount of evaluations needed for the complete picture (Factorial Design) may arise. Thus many schemes, such as Fractional Factorial Design, Box-Benkhen Design etc. [94] were developed to alleviate this problem. They trade off the ability to observe distinct effects (aliasing) for the reduced amount of evaluations needed. The increase in computational power over the last 40 years, has allowed scientists and engineers to use them as tools for modeling of (among others) physical systems. Thus, such modeling facilitated a new branch of Experimental Design called Virtual Experimental Design or Sampling. Here with the increase of computational power, allowing for multiple 'experiments' (i.e. numerical models) to be evaluated, also came a desire to increase precision and explanatory power. Thus, modern experimental design is a bit different from the classical one. The two key differences are:

- The amount of factors under consideration: now models that depend on a wide array of factors and factor combinations are present, which increases the design space.
- Continuous parameter spaces: now instead of discretizing the parameters into a set of 'high' and 'low' values it is often more common to apply a certain range or a statistical distribution for a given parameter and then

sample real valued samples from their distribution. This also increases the cost of experimental design.

This two key differences facilitated a development in, what is broadly called, Sampling Schemes. These schemes allow for efficient sampling of the (now highly dimensional and continuous) factor spaces with relatively good coverage of the space. One section of such techniques are the widely applied Low-Discrepancy Sequences.

3.1.1 Low-discrepancy Sequences

Discrepancy is a mathematical measure that serves as a foundation for a whole new branch of sampling methods. Discrepancy, D_N with respect to a sequence $(s_1, s_2...)$ and the interval $[a, b]$ and some sub-equisize interval $[c, d]$ is defined as:

$$D_N = sup_{a \leq c \leq d \leq b} \left| \frac{|s_1, ...s_N| \cap [c, d]}{N} - \frac{d-c}{b-a} \right| \tag{3.2}$$

Unpacking the definition one can see that it is the least upper bound on the absolute difference between the fraction of the set of points present within a chosen subinterval of the total sample size and the space that is occupied by the subinterval in the relation to the total size of the domain. A sequence would be equidistributed if discrepancy tends to zero and/or as N tends to infinity. In other words, discrepancy can be characterized as "lumpiness" of the sequence of points in a multidimensional space. The smaller the discrepancy, the less lumpy and more covered the design space is. Regular random sequences of k-dimensional points usually end up having very high discrepancy (e.g. lumpiness, like in Simple Random Samling), but there are infinite sequences of k-dimensional points that have low discrepancy, hence the name low-discrepancy sequence. [74]

One such sequence is the Halton Sequence. It works by first listing the amount of samples as indices in order corresponding to the natural numbers, i.e. starting from 1 and going to N (sample size). Then for each of the n factors, these indices are expressed in a increasing base system (i.e. first factor in base-2, second factor in base-3 etc.) where consecutive bases belong to the prime numbers. Then the numbers are reversed and placed after a decimal and then converted back to a decimal system. Samples from a Halton Sequence can be seen in Figure 3.1 which compares it against other techniques. Simple Random Sampling

(3.1, top left) is a basic sampling technique that has no structure imposed on it and simply uses the inverse transform sampling to produce new samples. This leads to production of excessive clusters that do not cover the sampling space properly. [71] A sampling scheme that imposes some structure on the sampling process is the Latin Hypercube Sampling (most widely applied method as per the Introduction), see 3.1 top right. Notwithstanding the imposed structure, this sampling still produces clusters. Low Discrepancy Sequences depicted in the lower half of Figure 3.1 produce less clusters than SRS or LHS. From the Figure it is readily observable that the low discrepancy sequences are better at covering the available experimental space than the LHS and especially better than the cluster-prone SRS. Selecting between sequences is difficult as only for very specific applications the differences come up, but a recommended technique for larger dimensional problems is the Hammersley sequence (it is an extension of the Halton sequence) and thus it is used throughout this work.

One of the properties of such sequences is that the estimates for various statistics, e.g. the mean, tend to more quickly approach the true values. For sequences the rate is of $\frac{lnN^n}{N}$ where N is the amount of samples and n is the amount of dimensions. This is generally faster than the rate of $\frac{1}{\sqrt{N}}$ for SRS and LHS. [74] This is a very useful property as when faced with estimating expectations (means) for Monte Carlo based Sensitivity Analysis, the convergence towards the true values is much faster. [74]

3.2 Gaussian Processes Regression

One of the two considered surrogates in this work is the Gaussian Process Regression. It is based on the idea of Gaussian Processes. A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution. [113] According to the definition, a Gaussian Process (GP) is a stochastic process, that is, a collection of Random Variables (RVs). Not just any random variables - normally distributed Random Variables. Such a collection is called a Stochastic Process, they can be finite and infinite. Gaussian Processes are typically defined as infinite collections of RVs. Most often, however, we desire to observe the behaviour of a GP on (arbitrarily large) set of discrete points. What allows one to jump between the possibly infinite definition and finite application is the Marginalization Property of Multivariate Gaussians. This propery can be mathematically shown for 2-D Gaussian as [114]:

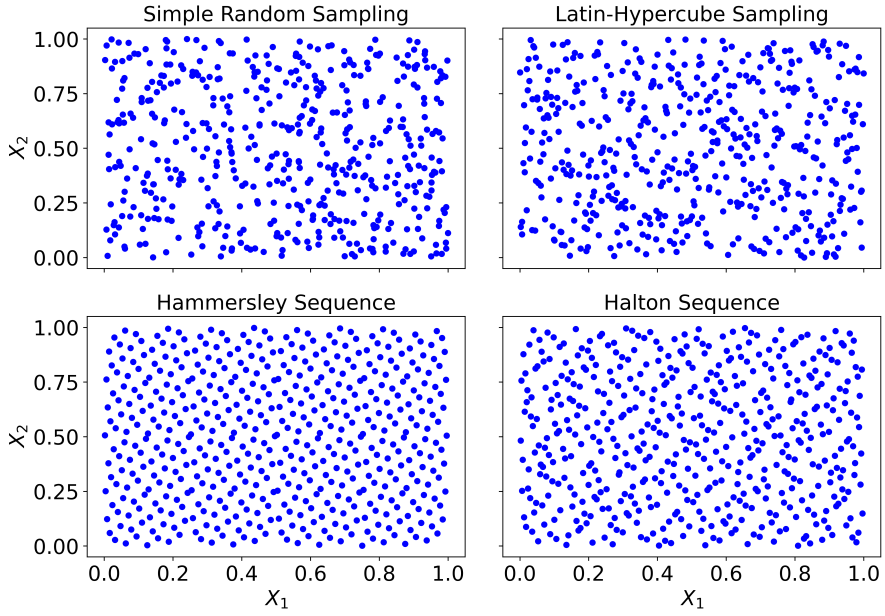

Figure 3.1: Comparison of the samples drawn from four different sampling schemes for a two-parameter design space, where both parameters are defined on an interval $X_i \in \langle 0, 1 \rangle$ and the sample size is large (N=500). Generated using artificial data.

$$p(X)dx = \int_y p_{X,Y}(x, y)dy = \int_y p(X|Y)p_Y(y)dy \tag{3.3}$$

The interpretation is the following: if we want to look at the smaller set of variables than before, the larger set need not impact the results we observe. In short it means we can "cherry pick" the set of RVs we care about and only examine their co-dependence. Given that a Gaussian Process is a collection of Normal RVs it inherits their property. Gaussian Process is specified by two parameters. The first one is the mean function, $m(x)$ and the second one is the covariance function, $k(x, x')$ [113]:

$$m(x) = \mathbb{E}[f(x)] \tag{3.4}$$

$$k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))] \tag{3.5}$$

Gaussian Process is then often written as:

$$f(x) \sim GP(m(x), k(x, x')) \tag{3.6}$$

The above line states that some (usually latent) function is distributed as a Guassian Process wit mean function $m(x)$ and a covariance function $k(x, x')$. Although Gaussian Processes themselves are interesting, the goal is applying them in regression context, thus Gaussian Process Regression (GPR) has been developed. There are a few key differences between standard, frequentist approach to regression and Gaussian Process Regression. The primary ones are:

1. GPR is a non-parametric regression model.
2. GPR is a Bayesian regression techniques

First of all GPRs are non-parametric. It means that unlike a typical SLR model (see section 2.4.3), where one tries to relate the predictors to the data via some parameters, Gaussian Processes essentially have 'infinite' parameters, i.e. as many parameters as datapoints or testing locations. This allows for the incredible flexibility of Gaussian Processes, however it comes at a cost. In parametric models, once the parameter values are learned, only their values need to be stored in computer's memory when one wants to evaluate the model. Unless one is dealing with (Deep) Neural Networks (which typically have hundreads, thousands or more parameters) it is a very convenient feature. However in non-parametric models, there is a need to have access to all data at evaluation time. [90] This is one of a few small hiccups to come regarding Gaussian Processes.

Secondly, GPR is a Bayesian regression technique. In Bayesian regression one typically expresses initial belief (prior) about various model parameters (see Section 3.5) and then the likelihood for the data generation (e.g. Gaussian, Poisson etc.). The goal is to obtain the posterior distribution for the parameters, however due to its analytical untractability one often resorts to sampling methods (such as Markov Chain Monte Carlo or Gibbs Sampling). [92]. The key difference between the common Bayesian framework and GPR is that in GPR one expresses uncertainty not about the model parameters, but about the function that generated the data itself. Such a method of expressing uncertainty has been commonly termed Latent Function Inference [115]. The idea is

that the process generating the data can be described by some function, $f(x)$, and based on the observed data we want to be able to emulate its behaviour. Similarly to linear regression (see section 2.4.3), one can express the observed data as:

$$y = f(x) + \epsilon \tag{3.7}$$

where $\epsilon \sim N(0, \sigma_n^2)$. Inference in Gaussian Process Regression works as follows. A joint GP prior is put on training and testing values of the latent functions, f and f_* respectively. The joint prior can be presented as (mean function is often assumed to be zero):

$$\begin{bmatrix} f \\ f_* \end{bmatrix} \sim N(0, \begin{bmatrix} K_{f,f} & K_{*,f} \\ F_{f,*} & K_{*,*} \end{bmatrix}) \tag{3.8}$$

where $K_{A,B}$ refers to a covariance between two random variables A and B. In this case $K_{f,f}$ and $K_{*,*}$ express the covariance between the training or the testing values of the latent function. $K_{*,f}$ describe the cross-covariance between the training and testing values. Subsequently, using the Bayes Rule (see section 3.5) one combines these Random Variables with observations to obtain a joint posterior distribution:

$$p(f, f_*|y) = \frac{p(f, f*)p(y|f)}{p(y)} \tag{3.9}$$

where the data likelihood can be written as:

$$p(y|f) = N(f, \sigma_{noise}^2 I) \tag{3.10}$$

Notice that in the likelihood, there is only explicit dependence on the underlying latent function f due to the way y has been defined. In the end, one wants to perform prediction at unseen locations, f_*. This is obtained by marginalization of the posterior distribution to obtain the Gaussian predictive distribution:

$$p(f_*|y) = \int p(f, f_*|y)df = \frac{1}{p(y)} \int p(y|f)p(f, f_*)df \tag{3.11}$$

This means that the predictive distribution is just a marginal distribution, obtained by normalization of the multiplication of the joint prior and the data likelihood. Due to the fact that the above formula contains only Gaussian distributions, an analytical solution can be obtained using some linear algebra operations on multivatiate Gaussians [90]. This leads to the analytical formula for the marginal posterior distribution for the test locations [113, 115]:

$$p(f_*|y) = N(\bar{f}_*, cov(f_*)) \tag{3.12}$$

where:

$$\bar{f}_* = K(f_*, f)[K(f, f) + \sigma^2_{noise} I]^{-1} y \tag{3.13}$$

and

$$cov(f_*) = K_{*,*} - K_{*,f}(K_{f,f} + \sigma^2_{noise} I)^{-1} K_{f,*} \tag{3.14}$$

Immediately both of the mentioned features of GPR come into play. The non-parametric nature of the GPR can be seen from equation 3.13, where the data, y, is directly involved into prediction of unseen cases. Secondly the use of prior, posterior and the likelihood show that GPR is an intrinsically Bayesian technique. Formula presented in equation 3.12 is used for carrrying out predictions in Gaussian Process Regression, it is a multivariate probability distribution on the test latent function conditioned on the observed data. [113]

Since Gaussian Processes are uniquely specified by the mean and covariance functions it is important to have a look at them. While the mean function does play a role, it is often assumed to be zero (a zero vector). It is done, because the mean structure of the latent function is often not known beforehand (unless it is a very specific case, e.g. tissue modeling see [59]). Moreover, the mean function is learned through the marginalization formulas (i.e. even though the explicit dependence on mean values for f and f_* are omitted, the mean is nevertheless updated with new data). In GPR the spotlight is definitely placed on the covariance functions. When one assumes a linear form of the covariance function, i.e.:

$$K(f, f) = K(f(X), f(X')) = K(X, X') = X \cdot X' + \sigma_0^2 \tag{3.15}$$

one obtains a fit very similar to what standard Bayesian Linear Regression would produce [113]. When one would like to reproduce more complicated functions, there is a need to introduce more complicated dependencies. This is done by introducing kernels, i.e. kernelizing the covariance.

3.2.1 Kernels in GPR

Kernels, $K(x, x')$ are generally functions that map a (often) vectorial set of inputs to a real number, i.e. $K : (x, x') \mapsto \mathbb{R}$. [116] One of the most widely used and 'boiler plate' kernels is the Squared Exponential (also called Radial Basis Function, RBF) kernel. It can be written as:

$$k(x, x') = \sigma_f^2 exp(-\frac{1}{2l^2}(x - x')^2) \tag{3.16}$$

where the difference, $r = (x - x')$ describes the distance between two different points (or vectors). It is sometimes called a proximity or similarity metric and that's why kernels are often denoted as similarity measures [116]. l is the kernel lengthscale and σ_f^2 is the kernel variance, these hyperparameters control the smoothness and and the amplitude, or range of values, produced by the kernel. The lengthscale itself describes how far does the kernel 'reach', i.e. small lengthscales will down-weigh points that are farther apart from our point of interest. The effects of varying the hyperparameters can be seen in a prepared example depicted in Fig 3.2. Figures 3.2 A and B show the effects of varying the kernel lengthscale while keeping the variance constant. On the other hand, Figures 3.2 C and D show the effects of changing the variance, while keeping the lengthscale constant. To avoid clutter both parameters are denoted by a placeholder parameter η and depending on the context it refers to lengthscale (LS.) or the variance (σ^2). Figures A and C show the situation where we look at the kernel output for $k(x, 0)$ while figures B and D show the resulting GP fit under constant lengthscale or variance. Looking at the figures, starting with the Figure 3.2 A, it is observed that as one keeps increasing the lengthscale, the reach of the kernel increases, i.e points farther down still have an impact on the kernel output. Figure 3.2 B shows how varying the lengthscale impacts the GP fit for some synthetic data. When the lentghscale is too short we essentially get a clump of Gaussians, each per one datapoint. When one increases the lengthscale, more and more neighbours of a given point are taken into consideration producing a smoother fit. When one goes too far however, the fit stops trying to capture the data's pattern. Similarly, Figure

3.2 D changes the variance. Larger variances result in larger kernel outputs for the same lengthscale. There, one can see that by increasing the variance it is possible to better capture the overall trend of the data, unless we go too far and get an overfit resulting in the curve going through each and every point. [113]

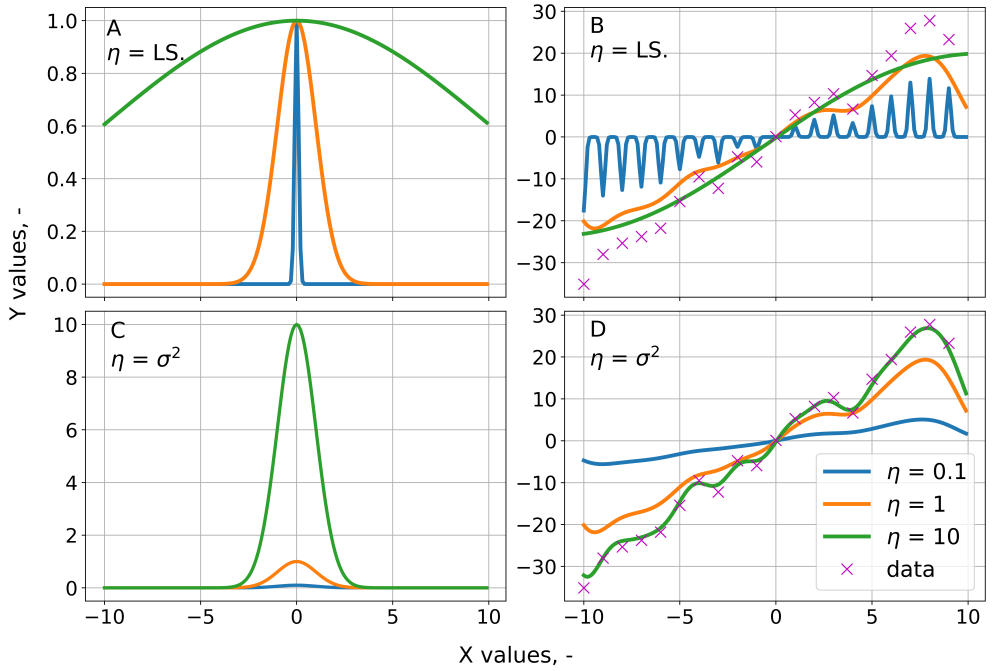

Figure 3.2: Effects of various hyperparameters on kernel and GPR behaviour. Figure shows the impact of varying parameters on kernel behaviour and resulting mean of the predictive posterior distribution. Impact of varying the lengthscale on kernel (**A**), impact of varying the lengthscale on the posterior mean (**B**), impact of changing the variance on kernel output (**C**) and impact of changing the variance on the posterior mean (**D**). η stands for a given parameter either lengthscale (LS.) or variance (σ^2).

It is useful to be aware of rules that kernels employed in GPR should follow. Since in GPR kernels are used to populate the covariance function, it is important that they follow its structure and properties. The most important properties are positive semi-definiteness and symmetry. These two properties arise from the structure of the covariance matrix, which the kernels are meant

to fill. [117, 67] Some of the commonly used kernels in GPR are [113, 118]:

1. RBF kernel
2. Matern32 or Matern52
3. Brownian kernel
4. Cosine kernel
5. Rational Quadratic
6. Polynomial Kernel
7. Bias
8. Linear Kernel

The influence of the kernel choice on the resulting Gaussian Process Regression fit is immense. It is easy to observe how the change of the kernels, that make up the covariance matrix, impacts the shapes of samples from the Multivariate Normal distribution depicted in Figure 3.3. There is no universal way to choose a kernel and the choice is highly dependent on the data at hand. Each of the kernels can contribute something to the overall shape of the function, e.g. the cosine kernel encoding the periodicity (e.g. in time series or signal processing applications), the linear kernel indicating a general linear trend (increase, decrease) and the RBF kernel encoding the local variation of the function [118].

There are also situations where any single kernel cannot properly describe the data, one of the possible solutions to such a problem is considering kernel combinations. Given that a new kernel that emerges from such combinations must preserve the properties of positive semi-definiteness and symmetry, the combination possibilities are somewhat restricted. The three possibilities for combinations of kernels, $k_1(x, x')$ and $k_2(x, x')$ for GPs include [113, 118]:

- summation:
$$k_1(x, x') + k_2(x, x') = K(x, x') \tag{3.17}$$
- multiplication:
$$k_1(x, x') \cdot k_2(x, x') = K(x, x') \tag{3.18}$$
- functional composition:
$$k_1(x, x') \star g(x) = K(x, x') \tag{3.19}$$

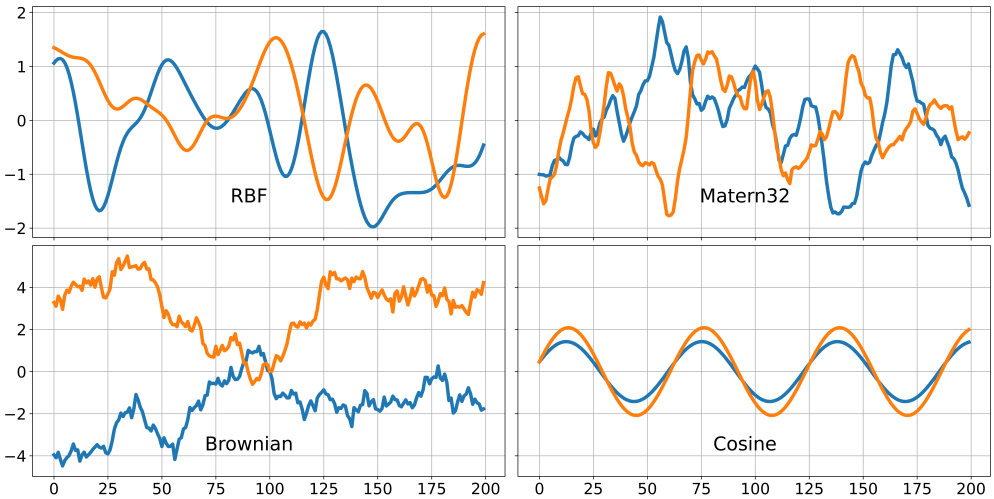

Figure 3.3: Samples drawn from a Multivariate Normal showing the impact of various kernels for covariance matrix on the sample shape.

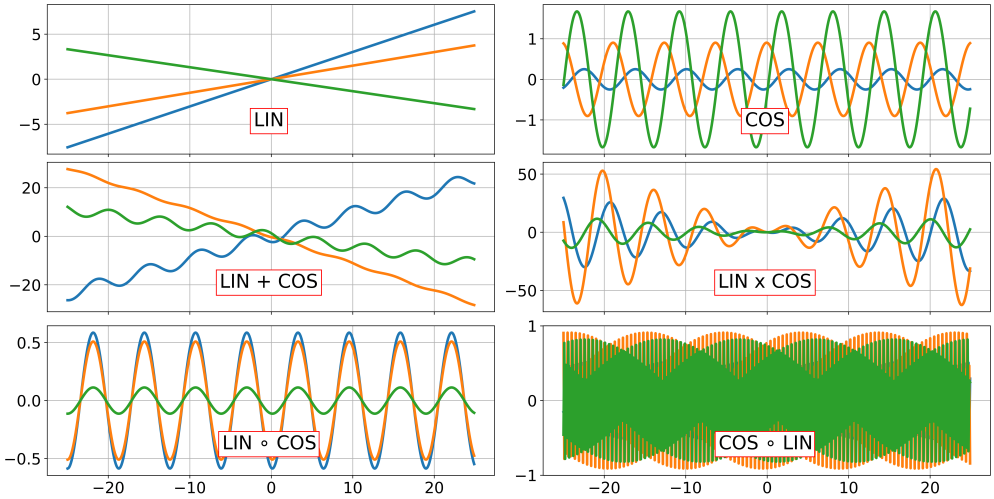

Figure 3.4: Effects of kernel combinations on resulting MVN samples. Top row shows the draws from MVN with a covariance specified by linear and cosine kernels respectively. Middle row shows the effect of adding and multiplying these kernels together. Bottom row demonstrates the effect of composing these two kernels with each other in different order.

where $g(x)$ is some function of x or another kernel. Each of the possible combinations yielding valid kernels can be seen in Figure 3.4. The plots in the top row correspond to MVN sample draws with a given covariance kernel. The middle row shows the impact on the samples drawn from the MVN under addition (middle row, left) and multiplication (middle row, right) of two kernels. For addition one can observe that the result is just a superposition of the shapes of two kernels, i.e. to a linear trend, periodic variation is added. For multiplication one can see that the linear signal attenuates the cosines when one approaches zero and then the values increase outwardly. Bottom row of the Figure shows the effect of the composition of two kernels. There the situation does not look so straightforward. Bottom left shows the composition of the linear kernel with the cosine kernel. Here one can observe that the trend of the linear kernel directly scales the cosine kernel output. For the last case, the composition order is reversed. This image further enforces the idea proposed by [118, 90, 115, 113] that kernel choice is highly subjective and is itself an art of sorts.

3.2.2 Hyperparameter Optimization

Once one sets up the kernels and Gaussian Process assumptions and update formulas, it may seem all the work is done. However as demonstrated in the previous section, varying hyperparameters has quite a big impact on the posterior distribution of the Gasussian Process. Since manual tuning is out of the question, closer attention needs to be paid towards doing it automatically. This is done through a process called hyperparameter optimization. This is done by optimizing the log marginal likelihood (also called the evidence). The likelihood is a function of hyperparameters, conditioned on data, and the maximization problem is set up the following way:

$$\theta^* = argmax_\theta \quad logp(\theta|y) \tag{3.20}$$

where $p(\theta|y)$ is the likelihood function

$$p(\theta|y) = N(y|0, \sigma_n^2 I + K_{f,f}) \tag{3.21}$$

since it is often assumed that $\mathbb{E}(f) = 0$. Upon plugging in the parameters to the normal distribution and taking the natural logarithm:

$$logp(\theta|f,y) = -\frac{1}{2}y^T K_y^{-1} y - \frac{1}{2}log|K_y| - \frac{n}{2}log2\pi \tag{3.22}$$

where

$$K_y = K_{f,f} + \sigma_n^2 I \tag{3.23}$$

is short-hand notation for the covariance for noisy targets. These terms in the log likelihood have readily interpretable terms [113]: the first term directly includes data thus is termed the data-fit term, the second one is the complexity penalty depending on the covariance matrix and the last one is a normalization constant. Optimization itself may be carried out using gradient-based algorithms, which should work well for simple cases. However when multiple kernels and/or kernel combinations are applied, the likelihood surface becomes highly nonlinear creating many possible local maxima and requiring a use of more involved optimization algorithms. [119]

The above formula for the marginal log likelihood shows the biggest problem with Gaussian Processes - scalability. Since there is an inverse of a matrix in the formula, most of the optimization algorithms will have to evaluate it (most often through Cholesky Decomposition), leading to a complexity of operations of order $O(N^3)$, where N is the amount of datapoints. Not only that, but the covariance matrix needs to be stored in memory, leading to memory scaling of $O(N^2)$. [120] This seems to be very problematic, however there are a lot of approximations that try to alleviate the problem. They often base on an approximation to the covariance matrix. Sparse Variational Gaussian Process is the method utilized in this work, for a detailed information on it, the reader is referred to [120] and [115]. It is based on the idea of introducing 'inducing variables', whose goal is to approximate that training dataset with fewer 'key' data points that best represent it. The more datapoints are chosen, the closer the datapoints represent the data, with the limit being, that as the amount of inducing variables approaches that training set size, it is wholly recreated. [120] Inducing variables are trained alongside other parameters in the model (like kernel hyperparameters) so that their representation of the dataset is optimal [121].

3.3 Data driven Reduced Order Modeling

Another avenue pursued in this work for model creation was the development of a Reduced Order Model. Generally there are two ways to construct a Reduced Order Model, i.e. intrusive and non-intrusive approaches. The intrusive approach relies on, for example, developing a POD-Galerkin projection that serves as an approximation to a given field quantity of interest (e.g. velocity) and then substituting the developed approximation straight into the Equations of Motion (e.g. Navier-Stokes equations). Although such an approach is feasible for Fluid-Structure Interactions [122], it is very difficult to implement and needs a lot of effort to reimplement when the model itself is subject to changes. It is especially difficult for FSI given that it is a coupling of two different fields (fluid and solid), which demands a large mathematical overhead. Non-intrusive approaches often rely on development of a Reduced Basis [123]. Such a basis is often obtained with a use of matrix decomposition techniques. There are many existing matrix decompositions (Cholesky, QR, etc.) that serve varied purposes. One that has enjoyed probably the widest application is the Singular Value Decomposition. It can be stated mathematically as [124]:

$$A = \begin{pmatrix} | & | & ... & | \\ a_1 & a_2 & ... & a_m \\ | & | & ... & | \end{pmatrix} = USV^* \tag{3.24}$$

where A can be some complex valued matrix, $A \in \mathcal{C}^{n \times m}$. n denotes here the number of rows in a matrix and m the number of columns. $U \in \mathcal{C}^{n \times n}$ are called the Left Singular Vectors, $V^* \in \mathcal{C}^{m \times m}$ are called the Right Singular Vectors. U and V^* are unitary (orthogonal when real-valued) matrices, and $(...)^*$ denotes the Hermitian conjugate (for real-valued matrices simply a transpose). $S \in \mathcal{R}^{n \times m}$ and is generally diagonal matrix, it contains the singular values on the principal diagonal. The reason for such a widespread usage of SVD in science in engineering are its properties and interpretation. Starting with the latter, U and V^* matrices encode the most important features in the matrix A (which most often is the dataset). U is in the column space of A, it is the new basis in which the entries of A are embedded. It contains all of the important patterns in the data, while V^* dictates what combination of U vectors is needed for reconstructing the matrix A with S dictating their relative importance. Thus, considering the present work, U would contain the time-dependent patterns in data, V^* would be responsible for putting them

together to reconstruct the parameter dependent (column-wise) representation in the data matrix (more on the problem set-up in section 5.4). Given that the SVD is a matrix decomposition, i.e. a matrix can be re-built from it, it is readily seen that the matrix S stores along its diagonal the relative importance (i.e. weights) of the singular vectors that are needed to reconstruct it. Thus it has been observed that it is possible approximate the given matrix with just a fraction of its original content by truncating the full decomposition and keeping just a small part of the original dataset. This leads to the following formulation (often called the 'economy' or 'truncated' SVD) [124]:

$$A \approx \hat{A} = \hat{U}\hat{S}V^* \tag{3.25}$$

where $\hat{U} \in \mathcal{C}^{n\times r}$ and $\hat{S} \in \mathcal{R}^{r\times m}$, where r is the truncation rank. Such an approach leads to a possibility of reducing and summarizing the original data by only losing a fraction of the accuracy. Not only the interpretation of the SVD, but maybe most primarily its properties have led to its wide adoption. First of all, by Eckart-Young theorem [124], it has been proven that given some rank-r trunctation of the SVD it is the optimal approximation to the matrix A in the least squares sense. Moreover, unlike many other decompositions, SVD is always guaranteed to exist (for any rectangular matrix) and the only hurdles in obtaining it may be its computation, where its time complexity tends to be $\mathcal{O}(nm^2)$ and space complexity is $\mathcal{O}(nm)$. In contrast to the Eigenvalue decomposition, the singular vectors unlike the eigenvectors are guaranteed to be unitary (orthogonal), which is important for creating an efficient reduced basis. Moreover, unlike eigenvalues, singular values are always real valued and positive. Among many other applications, SVD is used almost as a default for matrix inversion (Moore-Penrose Pseudoinverse) and thus in Least Squares. It is worth mentioning that SVD is also heavily related to Principal Component Analysis, a widely applied statistical method for reducing the order of the dataset by capturing the most pronounced correlations within the dataset. [124]

3.4 UQSA Theory

The broad strokes of UQSA have been described briefly in the Introduction. To recap, Uncertainty Quantification is an umbrella term that refers to various techniques applied to gauge the amount of uncertainty within a given model. It stems from the assumption that a given model is dependent upon parameters that are inherently uncertain, which leads to uncertainty present in the model

response [71]. It often takes the approach of calculating various statistical moments on the model response data to quantify how much variation is present. First of the two metrics, given some model response y, is the estimator of the mean model response, $\hat{\mu}_y$:

$$\hat{\mu}_y = \frac{1}{N}\Sigma_{i=0}^{N} y_i \tag{3.26}$$

and the estimator of the model response variance:

$$\hat{\sigma}_y^2 = \frac{1}{N-1}\Sigma_{i=0}^{N}(y_i - \hat{\mu}_y)^2 \tag{3.27}$$

Typically (however not necessarily), UQ precedes SA as it is generally a good practice to first know how much variance is present in the model before deciding to implement decisions based on SA's results.

3.4.1 Variance-Based Sensitivity Analysis

Variance-based methods have their roots in the High Dimensional Model Representation (also known as Hoeffding decomposition), which, assuming independent inputs, is written as [72]:

$$Y = f(X) = f_0 + \Sigma_i f_i + \Sigma_i \Sigma_{i<j} f_{ij} + ... + f_{12...k} \tag{3.28}$$

where $f_i = f(x_i)$, Y is random variable that is a function of other random variables, X_i. The decomposition has 2^k terms and is neither unique nor infinite. Moreover, Sobol proved that if each term in the decomposition has zero mean [74]:

$$\int f(x_i)dx_i = 0 \tag{3.29}$$

then all the terms of the decomposition are orthogonal pair-wise. The following dependencies based on output variance can be written: [125]

$$\begin{aligned} f_0 &= \mathbb{E}(Y) \\ f_i &= \mathbb{E}(Y|X_i) - \mathbb{E}(Y) \\ f_{ij} &= \mathbb{E}(Y|X_i, X_j) - f_i - f_j - \mathbb{E}(Y) \end{aligned} \tag{3.30}$$

By taking the variance of this decomposition one arrives at the ANOVA-HDMR decomposition: [74]

$$\mathbb{V}(Y) = \Sigma_i V_i + \Sigma_i \Sigma_{i \neq j} V_{ij} + ... + V_{12...k} \tag{3.31}$$

where k is the number of factors. Equation 3.31 fully decomposes the variance present in the response into many components, where $V_i = \mathbb{V}_{X_i}(\mathbb{E}_{\mathbf{X} \sim i}(Y|X_i)$ (and so on for higher order terms). This represents the variance across all X_i of the conditional expectation of the response, Y, conditioned at a given X_i, over all other variables than X_i, i.e. $\mathbf{X} \sim i$. This simply means that one takes a given value of the considered variable, X_i, and conditions on it the response subsequently averaging out all other variables. What follows is to calculate the variance over all possible values of X_i rather than just the one, to get V_i. By dividing the decomposition by the system response variance in Eq. 3.31, Sobol indices may be derived:

$$1 = \Sigma_i S_i + \Sigma_i \Sigma_{i \neq j} S_{ij} + ... + S_{12...k} \tag{3.32}$$

It can be seen that all Sobol indices must sum to 1. From the above decomposition the First Order Sobol index is described as [74]:

$$S_i = \frac{\mathbb{V}_{X_i}(\mathbb{E}_{\mathbf{X} \sim i}(Y|X_i)}{\mathbb{V}(Y)} \tag{3.33}$$

with the higher order ones following the same general formula. Sobol indices are simply a comparison of a given variance component (e.g. V_i) and the total response variance, $\mathbb{V}(Y)$ yielding a ratio indicating the relative 'importance' of each considered factor or factor combination. To see the proportion of the First Order effects contributed by variable, X_i, one refers to the First Order Sobol Index. It is most often used in factor prioritization setting, i.e. when one wants to seek out the most 'misbehaved' factor, variance-wise, and reduce its

assumed variability by further studying it. The Total Sobol index can readily be derived from Eve's Law (Law of Total Variance): [74]

$$\mathbb{E}_{X_i}(\mathbb{V}_{\mathbf{X}\sim i}(Y|X_i) + \mathbb{V}_{X_i}(\mathbb{E}_{\mathbf{X}\sim i}(Y|X_i) = \mathbb{V}(Y) \tag{3.34}$$

where, by rearranging, one may obtain:

$$S_{Ti} = 1 - \frac{\mathbb{V}_{X_i}(\mathbb{E}_{\mathbf{X}\sim i}(Y|X_i)}{\mathbb{V}(Y)} = \frac{\mathbb{E}_{X_i}(\mathbb{V}_{\mathbf{X}\sim i}(Y|X_i)}{\mathbb{V}(Y)} \tag{3.35}$$

Here, the numerator describes the conditional variance of the response, Y, conditioned on a given X_i, taken over all other variables than X_i, and then summed over all possible X_i's to condition on. This defines how much of the total variance can be attributed to a given factor, which includes all possible orders of variation and interaction. The Total Sobol index is commonly used in factor fixing setting, i.e. observing the total contribution of a factor to the total variance and deciding whether to downgrade it from a random variable to a constant value. [74]. It is possible to calculate Higher Order indices that yield more insight in variable interactions, however, this greatly increases the computational burden and tends to grow as 2^{k-1} (where k is the parameter count) when one wants to fully account for all possible interactions; it is further stipulated in [74, 72] that the first and total Sobol indices constitute a sufficient description of the model's variance.

3.4.2 Monte Carlo Methods for UQSA

In the ANOVA-HDMR decomposition, each variance component is described as an integral, i.e. by applying the fundamental relation $\mathbb{V}(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2$, one may obtain for V_i: [72]

$$V_i = \mathbb{V}_{\mathbf{X}_i}(\mathbb{E}_{\mathbf{X}_{\sim i}}(Y|X_i)) = \int \mathbb{E}^2_{\mathbf{X}_{\sim i}}(Y|X_i)dX_i - (\int \mathbb{E}_{\mathbf{X}_{\sim i}}(Y|X_i)dX_i)^2 \tag{3.36}$$

while the second integral is simply equal to f_0^2, the first one needs to be estimated. The most widely used approach relies on Monte Carlo approximation of the integral. However useful, Monte Carlo requires a lot of model evaluations and is only applicable either for simple models or for surrogates of the

Full Order Models. The latter setting is pursued within this work. The general framework for Monte Carlo estimation of the integrals relies on first splitting the evaluations of the model, Y, into 4 matrices: A, B, A_B^i and B_A^i. The A matrix is built out of the first half of model evaluations, while matrix B is built out of another half. The cross-matrices A_B^i and B_A^i are constructed by replacing the i-th column of a given matrix, with a corresponding column for the other one. Then the model is evaluated using these respective matrices, giving e.g. $y_A = f(A)$. [74, 72, 126] In this work, both Total and First order sensitivity indices are computed using three methods. For First Order indices:

$$\begin{aligned} S_i^{Sobol} &= \frac{\frac{1}{N}\Sigma(y_A \cdot y_{B_A^i}) - f_0^2}{\mathbb{V}(Y)} && \text{(Sobol [71])} \\ S_i^{Saltelli} &= \frac{\Sigma \frac{1}{N} y_B \cdot (y_{A_B^i} - y_A)}{\mathbb{V}(Y)} && \text{(Saltelli [72])} \\ S_i^{Jansen} &= 1 - \frac{1}{2N} \frac{(y_B - y_{A_B^i})^T \cdot (y_B - y_{A_B^i})}{\mathbb{V}(Y)} && \text{(Jansen [127])} \end{aligned} \tag{3.37}$$

and for Total Indices:

$$\begin{aligned} S_T^{Homma} &= 1 - \frac{(y_B \cdot y_{B_A^i}) - f_0^2}{\mathbb{V}(Y)} && \text{(Homma [128])} \\ S_T^{Sobol} &= \frac{y_A \cdot (y_A - y_{A_B^i})}{\mathbb{V}(Y)} && \text{(Sobol [129])} \\ S_T^{Jansen} &= \frac{\frac{1}{2N}(y_A - y_{A_B^i})^T \cdot (y_A - y_{A_B^i})}{\mathbb{V}(Y)} && \text{(Jansen [127])} \end{aligned} \tag{3.38}$$

where here the $\cdot$ represents the dot product and $f_0^2 = (\frac{1}{N}[y_A, y_B])^2$, i.e. basing on both inputs from A and B matrices to increase precision [72].

3.5 Bayesian Statistics and Estimation

First it is crucial to remind the Bayes Theorem, which was developed by its namesake, Thomas Bayes in the second half of the XVIII century and published

posthumously [130]. Currently it constitutes one of the fundamental results of probability theory. It is a simple mathematical statement relating the probability of one event, X, happening given some other event (e.g. data) y has already happened:

$$P(X|y) = \frac{P(X \bigcap y)}{P(y)} \tag{3.39}$$

where $P(X \bigcap y)$ is the joint probability of two events happening simultaneously. Through another fundamental rule from probability, i.e. Multiplication Rule, $P(X \cap y)$, can be rewritten as ([131]):

$$P(X \cap y) = P(y|X)P(X) \tag{3.40}$$

where $P(y|X)$ is the conditional probability of y on X and $P(X)$ is the marginal probability of X happening. Using the above two formulas and employing the Law of Total Probability, one may write: [131]

$$P(X_i|y) = \frac{P(y|X_i)P(X_i)}{\Sigma_{i=0}^{N} P(y|X_i)P(X_i)} \tag{3.41}$$

where $P(X_i|y)$ is the posterior distribution of X_i given y, $P(y|X_i)$ is the likelihood of the y given X_i, $P(X_i)$ is the prior (marginal) distribution on X_i and the denominator is simply a summation of all the possible states that X_i can take. This discrete Bayes Rule may also be rewritten in continuous form [92]:

$$p(X_i|y) = \frac{p(y|X)p(X)}{\int ... \int p(y|X_1, X_2, ...X_N)p(X_1, X_2, ..X_N)dX_1dX_2...dX_N} \tag{3.42}$$

The integral in the numerator is what immediately captures the attention: it is multidimensional and contains multiple complicated (and often impossible to derive) conditional probability distributions. Before computational advances of the XX century, it was this that kept Bayesian inference limited to simple problems, with conjugate prior-posterior pairs (given a certain likelihood) that produce neat analytical formulas for posterior probability calculation for tractable problems [132] or grid approximation techniques whose need for discretizing the probability space produced an exponentially growing problem,

computationally-wise. [132] The, already existing in mid-XX century, sampling based approaches, which sidestep this integral altogether were similarly or even more computationally prohibitive. Thus with more computational power, Bayesian approaches came more into prominence.

Bayesian statistics contrasts the classical approach of frequentist statistics used thus far. At the crux of the Bayesian-frequentist dichotomy lies the interpretation of probability itself. In frequentist approach the probability is viewed strictly as the relative frequency of an event occurring over an extensive series of hypothetical (not necessarily possible or real) repetitions of an experiment under identical conditions. Conversely, Bayesian statistics adopts a more subjective interpretation of probability. Within the Bayesian framework, probability signifies a measure of belief or uncertainty regarding the likelihood of an event, conditioned on available evidence or prior information.

Another fundamental distinction between Bayesian and frequentist statistics pertains to the treatment of uncertainty quantification. Frequentist methodologies typically employ confidence intervals and hypothesis tests to quantify and address uncertainty. They rely on the concept of a 'sampling distribution' which in the rawest sense is an imaginary distribution of a given parameter upon which conclusions and values for confidence intervals and p-values (and thus the results of hypothesis tests) are drawn. [92] While mathematically proven and widely applied, sampling distributions are not very convincing, especially to non-statisticians. Although there are many alternatives (e.g. bootstrapping [133]), Bayesian statistics provides one with an actual distribution, instead of an imaginary or assumed one. Bayesian modeling proposes an alternative and intuitive solution: draw samples from the actual distributions themselves, conditioned on the data observed and in turn provide distributions on each of the considered parameters. So unlike the classical approach of the sampling distribution, it draws samples from an actual distribution instead of just assuming one exists. An obvious elephant in the room is the fact that in Bayesian modeling, one needs to assume a lot of distributions - data likelihoods, priors (possibly hyperpriors in hierarchical cases). However all the other approaches (i.e. classical frequentist approaches or bootstrapping) also rely on (often implied or skipped over) assumptions of normality or some other distribution to facilitate inference (for non-normal frequentist approaches see robust statistics, e.g. [134]). While Bayesian approaches seem like it should be the 'default' method for data analysis, things like mathematical complexity, computational costs, extreme judgement when assuming distributions or limited software availability

have contributed to them being less popular than the frequentist method. [92]

3.5.1 Bayesian metrics for uncertainty

The shift in the approach to handling the uncertainty also necessitates a shift in the applied metrics and interpretations. The Confidence Interval common in Frequentist analyses, in Bayesian ones is reinterpreted to be based on the actual, and not sampling, distribution. It is termed a Credible Interval and has a much more straightforward interpretation of highlighting a range of values within which the true parameter value lies with a specified subjective probability, given both the observed data and prior beliefs. Again, unlike confidence intervals, which hinge solely on sample data, credible intervals combine prior knowledge with observed evidence, yielding a nuanced representation of parameter uncertainty. Another common Bayesian approach for quantifying uncertainty is the Highest Density Interval (HDI). The HDI is defined as the narrowest interval containing a specified proportion of the posterior probability density, i.e.:

$$HDI(\alpha\%) =< F^{-1}_{x,lower}, F^{-1}_{x,upper} > \tag{3.43}$$

where $F^{-1}_{x,lower}$ is the inverse Cumulative Distribution Function value of the sampled posterior density that constitutes the lower bound of the $\alpha\%$ HDI. Thus, from the definition, the HDI captures the most credible range of values for a considered parameter with a specified probability density, the most common ones cover the 90%, 95% and 99% of most probable parameter values and provides the lower and upper bound of said interval. This interval is characterized by its highest density region, distinguishing it from confidence intervals, which rely on specific quantiles of a sampling distribution.

While Bayesian analyses often result in full probability distributions, what often is required is some representative value of the whole sampled set. Thus, another common Bayesian term that is widely used is the Maximum A Posteriori estimate (MAP). MAP estimation is a method used in Bayesian statistics to estimate the parameters of a statistical model. It seeks to find the parameter values that maximize the posterior probability distribution given observed data. It can be expressed as:

$$\hat{\theta}_{MAP} = argmax_{\theta} \quad P(\theta|y) \tag{3.44}$$

In simpler terms it can be termed the most 'likely' value for a given parameter given the data. MAP can be linked to a Maximum Likelihood Estimation. MLE does not use any prior information and thus it is equivalent to performing a MAP estimation with a uniform prior - in this situation the likelihood is the most deciding factor in the MAP final estimate value, as it is in the MLE. However by incorporating some non-uniform and possibly informative priors, the equivalence between the two methods is only 'conceptual' and the superiority of the MAP can be easily seen. [92]

Another fairly unique technique in Bayesian modeling are the prior and posterior predictive checks. Since in this framework one must assume a lot of distributions and their respective parameters it is often a good idea to do a prior predictive check, which basically makes sure that the assumed quantities are a reflection of the analysts' intention. It proceeds by generating (hyper)prior parameter values that are the propagated down the model until a synthetic datapoint is obtained. A posterior predictive check works in a similar manner, the difference is that it is utilized after the model had the chance of observing data and drawing posterior samples. Thus both methods are useful - prior check allows to validate the assumptions made by the analyst and posterior check allows to test how well the obtained distribution would fare at generating future, unobserved data (they are often compared with a test set withheld before the analysis). [92]

3.6 Monte Carlo Methods for Bayesian Statistics

Markov Chain Monte Carlo (MCMC from now on) is a set of algorithms that relies on Markov Chains to converge to some given target distribution from which it aims at drawing representative samples. The sampling process itself is the 'Monte Carlo' part and its fundamental principle basically revolves around the idea that with infinite amount of samples from some target distribution, the distributional estimates under consideration tend to infinite precision. As mentioned, this method has been developed in mid XX-th century by Metropolis and Ulam [135]. It is a very common technique in statistics but also in numerical integration and other areas of engineering and thus won't be explained any further. Markov Chains on the other hand are far less popular in engineering circles and thus some background is required. Markov Chain is basically a connected progression of states (or random variables) θ_t that are linked together by some transition probability distribution, $T(\theta_t|\theta_{t-1})$. The fundamental property

of a Markov Chain is that the current state is only dependent on the previous one. While the proof of convergence of the Markov Chain to the target distribution is beyond the scope of this work (see [77, 131]), it is worthwhile to have a look at some properties a Markov Chain must satisfy to enable convergence. The properties are: [70]

- Aperiodicity
- Irreducibiility
- Stationarity

The simplest one to explain is aperiodicity. It simply means that there should exist no fixed period after which the chain repeats itself, i.e. returns to some state which it already has visited. Irreducibility means that a chain must be able to reach any existing state within a finite amount of steps. It basically states that there should not exist some regions in the set of all possible states that become impossible to reach. This ensures that the Markov Chain will be, in fact, albe to possiibly explore all avalilable and defined states and not produce biased estimates [131]. The final property is chain stationarity. This property is known as the Law of Detailed balance: [70]

$$p(\theta_t)T(\theta_{t+1}|\theta_t) = p(\theta_{t+1})T(\theta_t|\theta_{t+1}) \tag{3.45}$$

where $p(\theta_t)$ is the probability of visiting some state t. If there exists such a probability distribution $p(...)$ that satisfies the above equation it means that a Markov Chain is stationary and it is guaranteed that at some point it will converge to some stationary distribution, which allows one to trust that after some (however large) time, the stationary posterior distribution will be reached and sampled from.

3.6.1 Metropolis-Hastings and Hamiltonian Monte Carlo

The goal of Bayesian modeling is to obtain a posterior distribution over parameters of interest upon seeing and analyzing some data. In applied Bayesian statistics one obtains the posterior distribution directly from the Bayes Theorem, as in section 3.5:

$$P(\theta|y) = \frac{P(y|\theta) \cdot P(\theta)}{P(y)} \tag{3.46}$$

where previously X was a dummy variable denoting some event, now to align with most statistical literature it has been replaced by θ. It denotes a given parameter on which one wishes to perform inference. This formula is naturally extended to the continuous version [92]:

$$p(\theta_1, \theta_2, ...\theta_m|y) = \frac{p(y|\theta_1, \theta_2, ...\theta_m)p(\theta_1, \theta_2, ...\theta_m)}{\int p(y|\theta_1, \theta_2, ...\theta_m)p(\theta_1, \theta_2, ...\theta_m)d\theta} \tag{3.47}$$

where θ_i denotes the i-th parameter, y denotes the observed data, and $p(...|...)$ denotes a conditional continuous probability distribution. As mentioned before, the integral in the denominator is obviously problematic. It is common thus to approximate the posterior by removing the integral altogether and setting the posterior distribution proportional to the prior distribution multiplied by the likelihood: [92]

$$P(\theta|y) \propto P(y|\theta) \cdot P(\theta) \tag{3.48}$$

While it is possible to estimate the distributional parameters using Variational Inference, a more commonly applied method is Markov Chain Monte Carlo (MCMC) Metropolis-Hastings algorithm. Extensive books and articles have been written about the method (including seminal works of Metropolis [136] and Hastings [137] themselves), thus here only a short discussion of the method is provided. Having met the conditions of irreducibility, aperiodicity, and stationarity [77], a given Markov Chain is guaranteed to eventually converge to the target, and stationary distribution of states (random variables), the posterior distribution, given observed data y. MCMC M-H algorithm is one of the most widely used in Bayesian inference and it starts from some initial guess, θ_0. The algorithm then proposes a new parameter value with probability $q(\theta_1|\theta_0) = N(\theta_0, \sigma_{prop})$, where σ_{prop} is the variance of the proposal distribution. This proposal is used by the Metropolis version, however the extension made by Hastings generalized the proposal to any valid probability distribution (including unsymmetrical, multimodal etc.), $\pi(\theta_i|\eta)$, where η is some specific parameter set. Then for the M-H version, the algorithm evaluates whether to accept or remain at a current state (thus counting the current sample again) with the following decision rule [77]:

$$P(\theta_t, \theta_{t+1}) = \min\{1, \frac{p(\theta_{t+1}|y)}{p(\theta_t|y)} \frac{q(\theta_{t+1}|\theta_t; \eta_t)}{q(\theta_t|\theta_{t+1}; \eta_{t+1})}\} \tag{3.49}$$

where the first fraction is the ratio of the probability of the proposed sample given the current state of the posterior distribution to the previous one. The second fraction represents the ratio of the probability the current sample would come from a proposal distribution given that it was parameterized by some parameters, η_t, and the previous sample value. The denominator states the reverse. When a symmetrical proposal distribution is considered and it does not change over chain progression, the second fraction reduces to 1 and one is back at the classic Metropolis decision rule and algorithm. The algorithm then repeats the loop of proposal - decision rule - acceptance (or refusal) ad infinitum. It is important to note that when a given proposal is rejected, the current state (i.e. previous proposal) is 'accepted' or counted again.

While MCMC M-H is a widely applied algorithm, its biggest flaw is its inefficiency in the exploration of the parameter space (i.e. everything that happens before the decision rule). [138] To circumvent this shortcoming, Hamiltonian Monte Carlo was introduced. It proceeds by introducing an auxiliary artificial variable, referred to as 'momentum', r. This momentum variable is given its own distribution, often $r \sim N(0, M)$ where M is the 'mass' (correlation) matrix. Statistically, the joint posterior distribution can be written as:

$$p(\theta, r|y) = p(\theta|r, y)p(r) = p(\theta|y)p(r) \tag{3.50}$$

where $p(\theta|y)$ is the posterior distribution of θ and $p(r)$ is the marginal distribution of r. It is then assumed that the system's 'potential energy' (quotes apply as the energy in question is used solely by the means of an analogy rather than actual physical quantity) can be expressed as: [139]

$$E(\theta) = -\log(p(\theta|y)) \tag{3.51}$$

and the 'kinetic energy' may in turn be written as: [139]

$$K(r) = \frac{r^T M^{-1} r}{2} \tag{3.52}$$

By summing them together, one obtains a Hamiltonian: [138]

$$H(r, \theta) = E(\theta) + K(r) \tag{3.53}$$

which satisfies Hamilton's equations:

$$\frac{d\theta_i}{dt} = \frac{\partial H}{\partial r_i}; \frac{dr_i}{dt} = -\frac{\partial H}{\partial \theta_i} \tag{3.54}$$

that leave the Hamiltonian invariant over time. The Hamiltonian is then integrated with respect to time using a simplectic integrator to yield the phase space trajectories. After completing such an integration step, a sample is proposed and is either rejected or accepted. The acceptance criterion is: [138]

$$P(\theta_{t+1}|\theta_t) = \min\{1, \exp(H(\theta_t, r_t|y) - H(\theta_{t+1}, r_{t+1}|y)\} \tag{3.55}$$

which means that if the intergrator does a good job (i.e. time step is small enough and an appropriate number of steps is taken) the value of the Hamiltonian at the proposed parameter set will be very similar, $|H(\theta_t, r_t|y) - H(\theta_{t+1}, r_{t+1}|y)| \rightarrow 0$ giving a very high acceptance probability. While for MCMC M-H a typical acceptance is about $\sim 50\%$ [92], for HMC it is straightforward to get acceptance of 65% and above. [138] It is however not rare to see HMC and MCMC M-H implemented together, particularly in Hierarchical Models, where the parameter space is a bit more involved and the correlations between variables may impede the proper Hamiltonian parametrization [139].

3.6.2 MC-based model diagnostics

Naturally, as with every model, it is imperative to test its quality. Bayesian diagnostics are somewhat different from the frequentist ones and thus necessitate an introduction. There are many commonly employed diagnostics, but the three most widely used (and most informative ones) are the Effective Sample Size, R-hat (or $\hat{R}$ also known as potential scale reduction factor), and Monte Carlo Standard Deviation. R-hat starts by defining the within (W) and between chain variances (B):[77]

$$\begin{aligned} W &= \frac{1}{m}\frac{1}{(n-1)}\Sigma_j^m \Sigma_i^n (\theta_{ij} - \hat{\mu}_j)^2 \\ B &= \frac{n}{(m-1)}\Sigma_{j=1}^m (\hat{\mu}_j - \hat{M}^2) \end{aligned} \tag{3.56}$$

where m refers to chain count, n sample count, $\hat{\mu}_j$ is the sample mean and M is the grand mean (i.e. sample mean averaged across chains). Total variance is the estimated as the weighted combination of the variances: [77, 140]

$$\mathbb{V}^+[\theta|y] = \frac{n-1}{n}W + \frac{1}{n}B \tag{3.57}$$

where $\mathbb{V}^+[\theta|y]$ tends to overestimate the variance when the chains are unconverged and W tends to underestimate the Within sample variance for finite n. $\hat{R}$ has been defined to be: [77, 140]

$$\hat{R} = \sqrt{\frac{\mathbb{V}^+[\theta|y]}{W}} \tag{3.58}$$

R-hat is a measure of how much the scale of M would reduce if the sampling were to run indefinitely. As sample size tends to infinity, $\hat{R}$ tends to 1, i.e. the Within to Between chain variance is identical, i.e. all chains have converged at a similar reasonable estimate. Another useful statistic is the Effective Sample Size (ESS), which is responsible for tracking how many actual independent samples are present within the chain. ESS measures the amount by which autocorrelation within the chains leads to an increase in the uncertainty of the estimates. It is defined as: [140]

$$ESS = \frac{M_s \cdot N_s}{1 + 2\Sigma_{t=1}^{2m+1}\hat{\rho}_t} \tag{3.59}$$

where M_s, N_s refer to chain and sample count respectively. $\hat{\rho}_t$ is the combined autocorrelation estimate, defined as: [140]

$$\hat{\rho}_t = 1 - \frac{W - \frac{1}{M}\Sigma_{m=1}^{M}\hat{\rho}_{m,t}}{\mathbb{V}^+[M|y]} \tag{3.60}$$

where $\hat{\rho}_{m,t}$ are the autocorrelation estimates at a given lag t from multiple considered chains $m \in (1, ..., M)$. In the formula one can also see the (introduced before) Within-sample variance estimate W and multi-chain variance estimate $\mathbb{V}^+[\theta|y]$. If the chains have not yet fully converged, the variance estimator $\mathbb{V}^+[\theta|y]$ will, again, tend to overestimate the variance, leading to an overestimate in autocorrelation and an underestimate of the ESS. [140] A low ESS is

considered to be below 100 samples. When the ESS is very low, especially when the amount of samples is large, it indicates some fundamental problem with the structure of the model which needs to be resolved. Monte Carlo Standard Error (MCSE) is simply a standard error applied to a given Monte Carlo Estimator, i.e.: [140]

$$SE(\hat{\theta}) = \frac{\hat{\sigma}_\theta}{\sqrt{ESS}} \tag{3.61}$$

where SE is the Standard Error on the parameter and $\hat{\sigma}_\theta$ is just the standard deviation of the pooled posterior draws of some parameter θ. The smaller the estimated standard error, the closer the estimate is expected to be to the true parameter value.

4

Statistical analysis and FSI validation

The purpose of this section is to go from the construction and design of the experimental facility, through statistical analysis of the data, to ultimately end up at the numerical model validation. As mentioned in the Introduction, construction of the experimental rig is a crucial part in the development of the new method for local arterial stiffness estimation. Rig's purpose, itself, is twofold. First, create a large amount of data that can be used for numerical model validation. Second, to establish the ultrasound as a reliable method for measuring arterial displacements, that can be later fed into the developed algorithm. Having described the rig's construction and the measurement procedure, the chapter moves on to rudimentary data analysis to get a 'feel' for the data, i.e. to understand the trends present within and process it. What follows is a development of a more involved statistical model to better model the uncertainty present in the data due to e.g. equipment noise. This whole part of the chapter is based on author's published work, [141]. The final part is the numerical model development and validation. The validated numerical model will be then used in the development of a statistical surrogate, used to enable tasks such as parameter estimation.

4.1 Description of the experimental rig

The goal for the experimental rig, was to faithfully simulate the behavior of the Left Common Carotid Artery (LCCA) through the use of an arterial phantom and to allow the direct monitoring of its deformation under simulated cardiac pressure and flow cycle. The measured volumetric flow rate was kept at around 0.5 l/min in order to approximate physiological conditions in the LCCA of an average adult male [142]. To obtain representative data while working with limited resources, experiments were planned over a series of four systole-to-diastole pressure ratios, i.e. **A)** 110/70 mmHg, **B)** 120/80 mmHg, **C)** 135/95 mmHg and **D)** 140/100 mmHg. These particular four ratios were selected as they cover a wide range of physiological values and scenarios. They are often classified as per the following: A - optimal, B - normal, C - high normal and D - Grade 1 hypertension [143]. Each of them were repeated 7 times, as this was the amount of experiments at which the phantom still retained its original properties and was not extensively fatigued. These ratios were obtained by employing the Harvard Apparatus Pulsatile Blood Pump designed for large animals [144]. The pump was used due to its ability to closely resemble the ventricular action of the human heart. The setting of pump's stroke volume was at 15 ml, the stroke rate at 60 rpm, and the systole / diastole flow ratio at 35%: 65%. The scheme of the experimental rig can be seen in Figure 4.1, and a photograph of the physical facility itself, together with description of the most crucial components, is presented in Figure 4.2. The operation of the rig can be thought of as starting with the water reservoir which stores the water used in the experiments, since it was the working fluid. Then it moves on to the aforementioned pump. The path then proceeds into the elastic hydraulic accumulator, whose role was to mimic the action of the aorta. After the accumulator, the path diverges into two: one loops back to the water tank. The other fork in the path leads straight into arterial phantom. There, the most important part of the facility is located: a glass aquarium that holds the phantom, which simulates the behaviour of the human body's internals. It was filled with water to allow a medium for Ultrasound Waves to propagate. Located at the aquarium are the most important facilities used in the measurements: flow meters, pressure transducers, high-speed cameras (later the ultrasound) and the backlights.

Starting off with the flowmeters, the ones used in the experiment were the Endress + Hauser Dosimag electromagnetic flow meters [145]. Eight pressure transducers (Harvard Apparatus Blood Pressure Transducers APT300) together with Compact Transducer Amplifiers (Harvard Apparatus) were in-

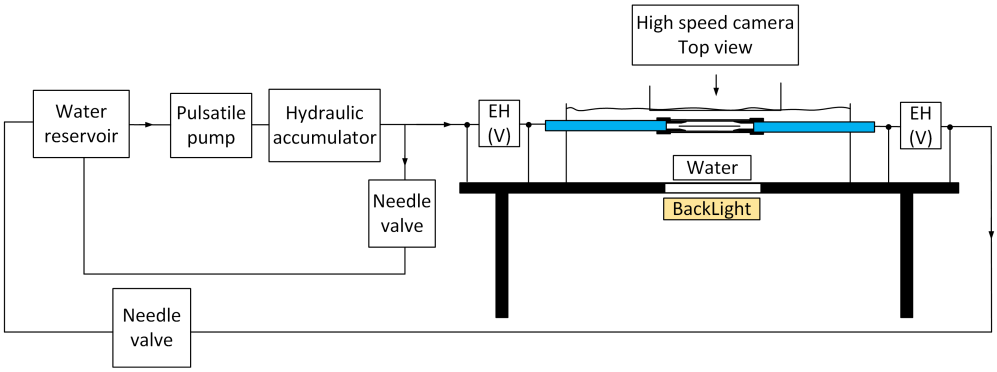

Figure 4.1: Schematic representation of the testing rig used for measurements. The scheme starts with the water reservoir from which water was drawn by the pulsatile pump. Then the fluid was passed straight to a hydraulic accumulator that mimicked the behaviour of the aorta. After this, the flowline was split into two: the first one back to the reservoir imitating the recirculation of the blood back to its source and to the carotid artery phantom, i.e. the testing location with high-speed cameras / ultrasound. The flowline then comes back to the reservoir creating a closed system. Figure reproduced from [141]

stalled to measure the pressure at several points proximal and distal to the phantom. To confirm that the displacement of the tube could be determined from the images of the phantom's deformation, the change in external diameter was monitored using two high-frequency cameras: a Phantom MIRO C110 and a Phantom VEO 710 [146]. The Phantom MIRO C110 recorded 1000 frames per second (FPS) at a resolution of 1280x900, which covered around 22% of the entire phantom's length. This was done using the irix lens [147] (150 mm f/2.8 Macro 1:1) combined with a 20 mm extension apparatus. The second camera, a Phantom VEO 710 was mounted with a NIKKOR 200 mm F/4.0 MACRO lens with another additional 68 mm extension apparatus. It recorded 1000 FPS of resolution 1280x800 while covering 30% of the phantom. The two cameras were mounted perpendicularly to each other, so that they capture the phantom's deformations in two separate, orthogonal directions (refer to Figure 4.2). Lastly, in order to eliminate background artifacts and sharpen the edges of the phantom in camera images, two backlights were installed.

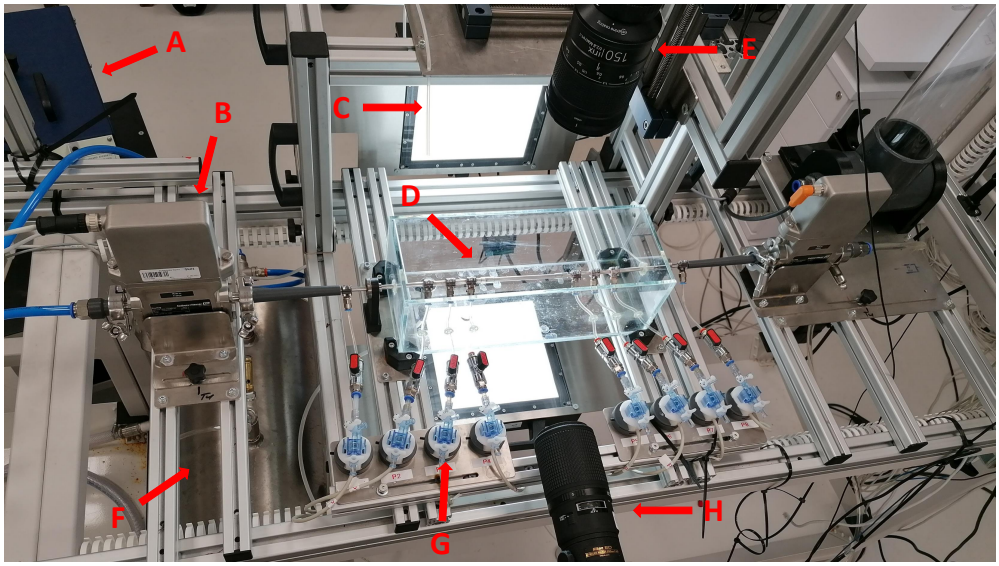

Figure 4.2: Photo of the configuration of the experimental rig, along with devices used for data acquisition. Arrows indicate selected components of the laboratory set-up. A - periodic pump, B - flowmeter, C - backlight, D - arterial phantom, E - top camera (MIRO), F - reservoir tank, G - pressure transducers, H - side camera (VEO). Reproduced from [141]

4.2 Measurement procedure

The data collection procedure implemented for the testing rig was controlled using an in-house program developed in LabVIEW (National Instruments Corp., USA) by Marek Rojczyk, PhD. This custom application allowed simultaneous monitoring and acquisition of pressure and flow data from measurement devices. These were gathered at 10 ms intervals (100 Hz), whereas image data was collected at 1 ms intervals (1000 Hz, i.e. 1000 frames per second). The high-speed camera recordings allowed precise tracking of the phantom's deformation. To facilitate simultaneous measurement of pressure and flow values every 10 ms, the field-programmable gate array (FPGA) implemented in the cRIO 9074 controller was used to read the values from the present measuring devices (8 pressure transducers and 2 flowmeters). Once the data are collected by the FPGA, each of the measured quantities was passed on to the real-time (RT) as a vector, through direct memory access (DMA FIFO). The RT loop's purpose is to collect considered vectors as one big matrix and, at 500ms in-

tervals, to send this matrix from RT to the application host (which here was the laboratory's PC). There the data is stored in a file. The whole process is depicted schematically in Figure 4.3. The Dosimag flowmeters were calibrated by the hardware suppliers, and subsequently checked by connecting the flowmeters in series with a turbine flowmeter (KOBOLD) [148]. Then, by imposing various volumetric flow rates, the read outs were compared against each other (and with the pump's setting) to verify their consistency. The calibration of pressure transducers and Compact Transducer Amplifiers (CTAs) was carried out by first isolating the phantom region of the facility, where the transducers were located, into a closed loop. To this closed system, a column of liquid was connected. Consulting the manufacturer's manual [149], two calibration points were selected. The first being 0 mmHg gauge pressure (equal to 0 Volt signal from the transducers). The second point was gauge pressure of 100 mmHg (equal to 1V signal, thus giving a 100 mmHg / 1 Volt ratio). The pressure was controlled by the height of a liquid column. At both endpoints (0 mmHg and 100 mmHg), the 'low' and 'high' values of the CTAs were adjusted to correctly reflect the applied pressure. To ensure synchronization of the recorded images with pressure and flow waveforms, a dedicated procedure was programmed into the control application.

Only after the flow and pressure measuring devices were tuned to the desired level, measurements could begin. The data recording was initiated manually in the LabVIEW application and the and the control system triggered the cameras to begin recording at detection of the first upcoming decreasing trend in volumetric flow. Both cameras (VEO, MIRO) then recorded images for 5 seconds total. This made it possible to record a couple of pump cycles, which in total produced around 5000 images (per camera). For each of the described systole/diastole pressure ratios, the procedure was repeated 7 times each (see the previous section).

4.3 Phantom Displacement Detection

In order to capture the displacements of the phantom from the recorded images, an automatic image-processing algorithm was developed using Python 3.9. Some of the applied libraries were: openCV [150], SciPy [151], Pandas[152] and NumPy [153]. The functioning of the algorithm is depicted in Figure 4.4. There, the left branch shows the way in which the lateral dimension ('diameter') of the phantom was determined from a recorded image. The right branch,

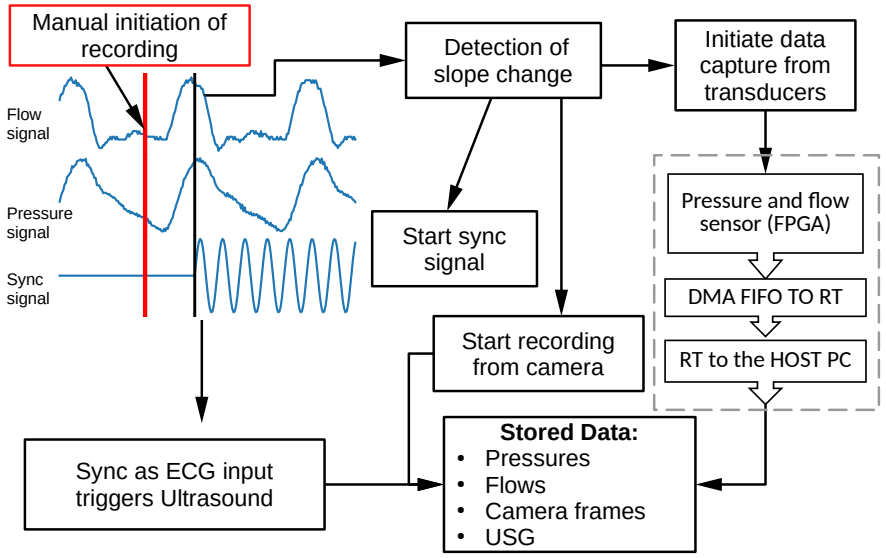

Figure 4.3: The start and synchronization of data acquisition across the equipment is depicted schematically. First, as the conditions stabilize, the system is triggered manually sending a signal (red vertical line). Then the signal from the flowmeter is closely monitored. Upon detecting a decreasing trend in flow (black vertical line), data collection begins with a synchronization signal being generated and sent to the ultrasound system. The inset with dashed lines illustrates the programming of the FPGA to read pressure and flow data every 10 ms. These values are received by an RT loop in the controller and stored in memory, and subsequently the values are read from this memory to the host PC where they are collected in a data file. (reproduced from [141])

on the other hand, shows the procedure for determination of pixel dimension for each of the cameras. The raw images from both cameras only record pixel information, thus to extract distances and displacements in millimeters, the linear dimensions of the pixel were determined for both cameras. As both cameras have isotropic pixel dimensions, it is sufficient to determine the dimensions along a single axis. A calibration procedure was carried out at the beginning of each measurement. First, a linear scale suited for microscopy (for high pre-

cision) was placed in the same position that the phantom would occupy, and each camera recorded an image of this scale. The images of the scale were filtered to detect edges (see the right branch of Figure 4.4). This resulted in a filtered image with a strip of points located on each vertical edge corresponding to the edges of the divisions of the linear scale. The thickness of a single vertical division of the scale is defined by the two edge points $X_{i,\text{left}}$ and $X_{i,\text{right}}$, which are averaged to obtain the center $X_i = (X_{i,\text{right}} - X_{i,\text{left}})/2$ (red dots in the right branch of Figure.4.4). The length, l_{true}, (in millimeters) between the left-most and right-most lines was determined by the division of the microscopic scale (0.5 mm) and the number of division lines captured in the image. The length in pixels, l_{pixels}, covered by the linear scale was calculated as the difference between the average pixel coordinates of the left-most division line and the average coordinates of the right-most division life. The linear dimension of the pixel was then obtained by dividing the calculated length, l_{true}, by the distance in pixels. Across all experiments and cases, the median scale for the MIRO camera was 112.51 (interquartile range 0.19) pixels/mm while for VEO the median scale was 76.75 (interquartile range 0.39) pixels/mm. To verify the calibration procedure, the linear dimensions were also determined with ImageJ (LOCI, University of Wisconsin, [154]) for several test cases and identical results were found.

Displacements were then determined from each image of the phantom recorded during each measurement. The first step was blurring of the image, followed by applying a Canny detection filter (see left branch of 4.4) with high and low thresholds of 100 and 200 to produce an image with pixel values of 0 everywhere except at the edges of the tube [84]. The distance between the edges was determined by traversing the image vertically at a given horizontal position, X, and storing (X, Y) coordinates of nonzero pixels (red dot in the left branch of Figure 4.4). For a given horizontal position X_1 there were two points, (X_1, Y_1) and (X_1, Y_2), corresponding to opposite vertical position on the tube (Y_1 and Y_2) at some fixed horizontal position (X_1). Next, the diameter in pixels, d, was calculated as the difference between these two Y coordinates. This process was repeated in the horizontal direction (positive X direction) and for each image yielding diameters at each horizontal location for each time, $d(X, t)$. Then a spatial average was computed $\bar{d}(t) = \frac{1}{N}\sum_X d(X, t)$. Finally, as the main quantity of interest is the displacement, the difference $d_{\text{rel}}(t) = \bar{d}(t) - \bar{d}_{\text{diastole}}$ is computed and analyzed in the remainder of the work. The diastolic diameter, $\bar{d}_{\text{diastole}}$ was defined as the minimum $\bar{d}(t)$ determined over the entire recording.

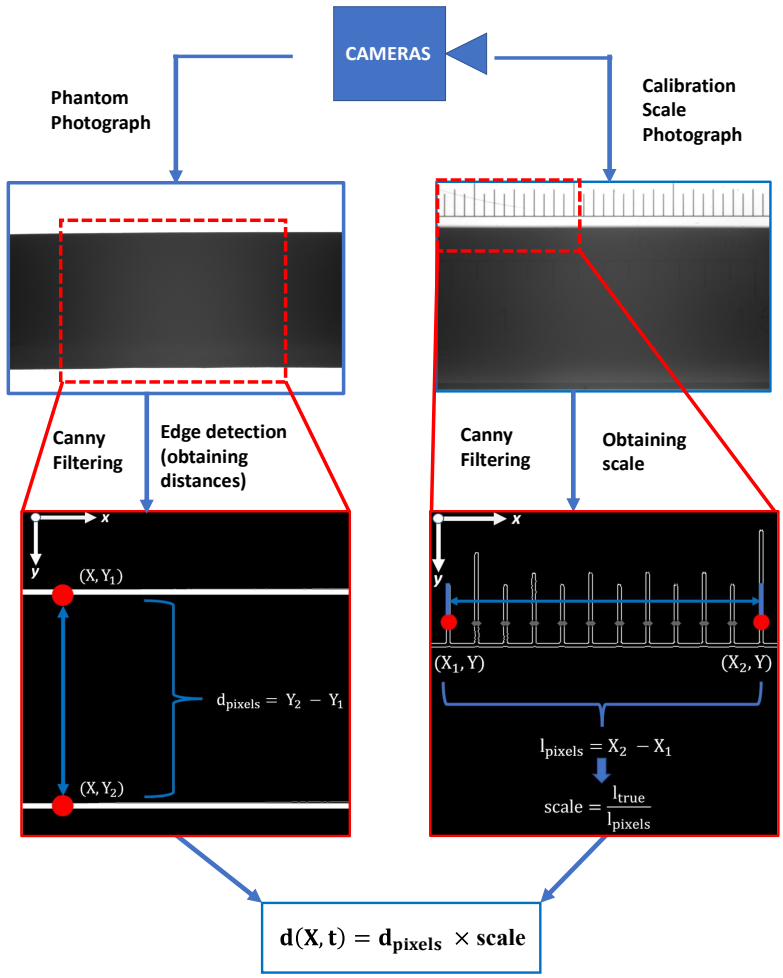

Figure 4.4: Image processing workflow. The left side depicts the determination of diameters from a filtered image. At each column of pixels, X location, the filtered image has two corresponding Y locations. The distance between them (after converting from pixels to millimeters) corresponds to the phantom diameter in pixels. The right side depicts the process of obtaining the conversion factor between pixels and millimeters. Filtering the image of the linear scale yields left and right edges $X_{i,\text{left}}$ and $X_{i,\text{right}}$ at each scale division. The positions of the first and last scale divisions are determined by averaging the position of the divisions' left and right edges, $X_i = \frac{X_{i,\text{right}} - X_{i,\text{left}}}{2}$. Then the distance is calculated as $l_{pixels} = X_2 - X_1$. Subsequently the pixel edge length (millimeters per pixel) was obtained by dividing the true distance by the pixel distance. Finally, the phantom diameter in millimeters is computed given the diameter in pixels and the conversion factor. Reproduced from [141]

4.4 Pressure And Flow Data Processing

It is often the case that the data resulting from measurements possess some noise - i.e. something that is a part of the signal but does not reflect its underlying structure. Although the noise is generally unwanted, some of it can be often easily handled and interpreted. However, it becomes a burden when the noise is so large that it produces outliers. Outliers are defined as unrepresentative datapoints deviating very strongly from the bulk of the data - there are many possible sources for outliers, e.g. they may be a result of some equipment malfunctions (e.g. in engineering measurements) or small sample size (e.g. in statistical polling). [67] In classical statistical approaches, outliers are often a point of debate - whether to remove them or not is often the primary question at hand, as any modeling or inference that includes them may produce heavily biased and unrepresentative results. However, in the context of this work, it is pretty clear what the pressure and flow curves should look like and any outstanding outliers are a result of some equipment malfunction or artifacts, rather that an actual sudden improbable spike in flow or pressure. In the case of pressure and flow data, as well as for image data, it is necessary to assess it's reproducibility so that it is clear whether the experiments produce desired results that can be reproduced by others and that they were actually repeatable. In order to assess these qualities, first the data measured needed to be processed correctly to remove any outliers and prepare them for a more advanced statistical analysis that could produce a set of summary statistics that would answer the reproducibility and repeatability question with actual, quantitative results (see section 4.5). In Figure 4.5 the raw (flow transformed to L/min from Hz and pressure to $mmHg$ from V) data from pressure and flow transducers are presented, next to each other. In the Figure, it is obvious that the data have some underlying noise present in them (small wiggles in the general trend) but also some outliers are noticeable, particularly in the flow data (see the red marker). This image reinforces the idea that outliers can be removed as the measured rapid increase in the flow value is not accompanied by any response from the measured pressure values, which would be generally expected.

After initial pre-processing with the Hampel and low-pass filters, the data were split into individual cycles in order to average each point of the cardiac cycle over time. This was achieved by employing a simple, developed minima-seeking algorithm to identify the start point of each cycle. Starting from the beginning of the data, the algorithm would seek out minima by moving through the pres-

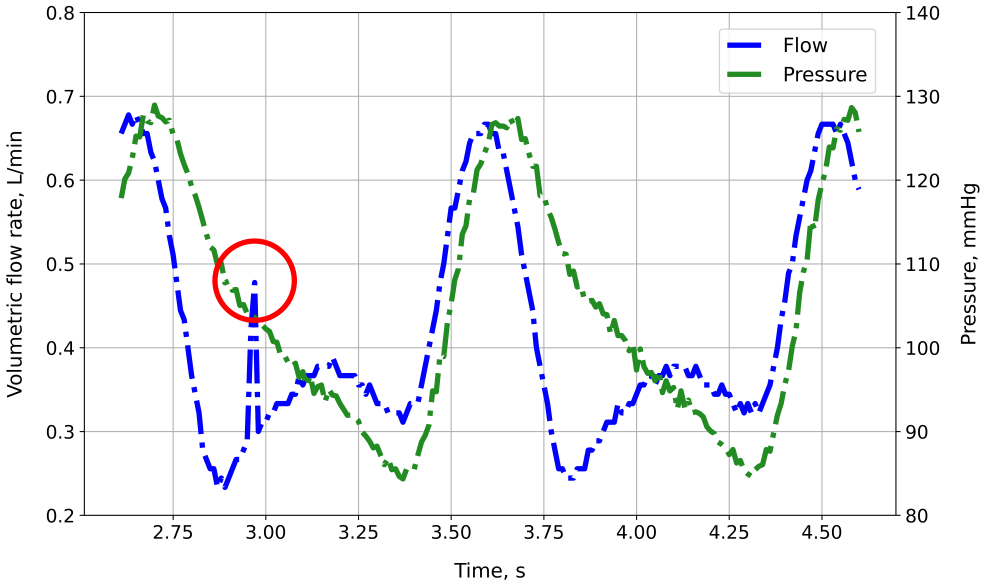

Figure 4.5: Flow (blue) and pressure (green) raw data plotted next to each other. Some underlying noise in the data can be seen and some outliers (red circle) are also present indicating the need for processing.

sure and displacement data with a fixed step, t_{step}, and then looking for the minimum in a fixed window, ($t_{step} - w$, $t_{step} + w$), where w was a constant determining window width and t_{step} is the expected cycle length. This algorithm was quite robust for pressure and displacement data (the determination of displacement data is described in Section 4.3), since their end-cycle minima were clearly isolated, it was insufficient for the flow data that exhibited many low amplitude oscillations at the tail end of each cycle. These impeded automatic determination of the minima corresponding to the end of the cycle. Since pressure and flow data were collected with the same frequency and were expected to have the same duration, power spectrum density [155] of the pressure data was computed to identify the dominant frequency corresponding to the cycle length. This frequency was then used to determine t_{step}, and the width of the window was reduced to avoid splitting cycles due to noise or flow oscillations near the end of the cycle.

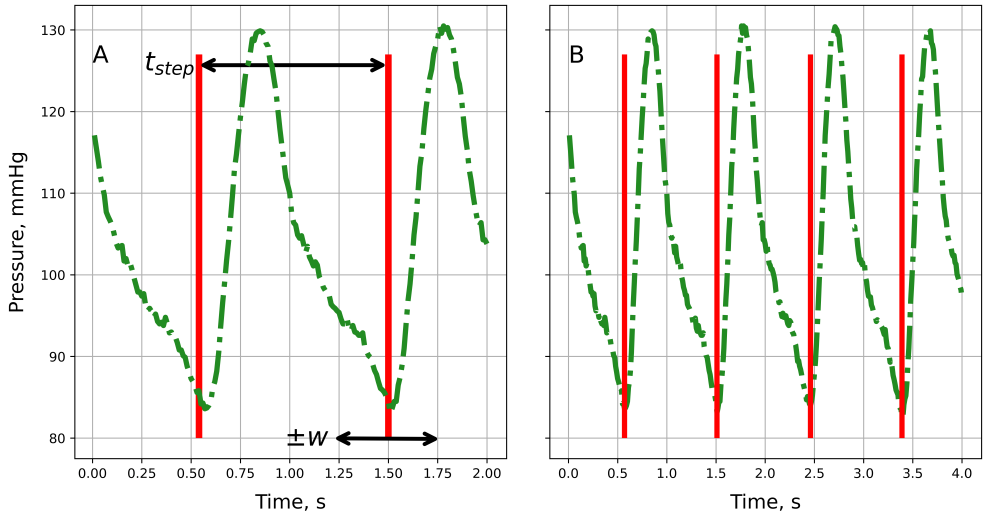

Figure 4.6: Illustration of the cycle splitting algorithm on pressure curves. **A**: The algorithm takes a step, t_{step} and the begins to look for a minimum within a fixed window of width $\pm w$, i.e. $2 \cdot w$. When a minimum is found, a split is made. **B**: Visualization of the splitting performs when working on the larger subset of the entire data series.

4.5 Modeling data with Linear Mixed-Effects

Equipped with all the statistical knowledge necessary from Chapter 2, one is able now to appropriately model the measured data. As described in Section 4.1, the experimental procedure was repeated 7 times for each of the four systole/diastole pressure ratios, and a simple preliminary statistical procedure was carried out to characterize the variability of the pressure and flow waveforms. Figures 4.7 and 4.8 show the across-time and across-measurement averaged waveforms for both pressure and flow data respectively. In both figures a 95% confidence interval [91] was computed by:

$$\mu \in \bar{x} \pm t_{\alpha/2,n-1} \frac{S^2}{\sqrt{n}} \tag{4.1}$$

where μ is the population mean (true value) for pressure (or flow) values, $\bar{x}$ is the (sample) data mean first calculated across measurement series and then within each averaged measurement, S^2 is the (sample) data variance, $t_{\alpha/2,n-1}$ is the critical value for a two-sided confidence interval based on Student's t-

distribution and n is the amount of cycles within a measurement. The resulting confidence intervals for the average waveforms suggest that both pressures and flows were repeatable.

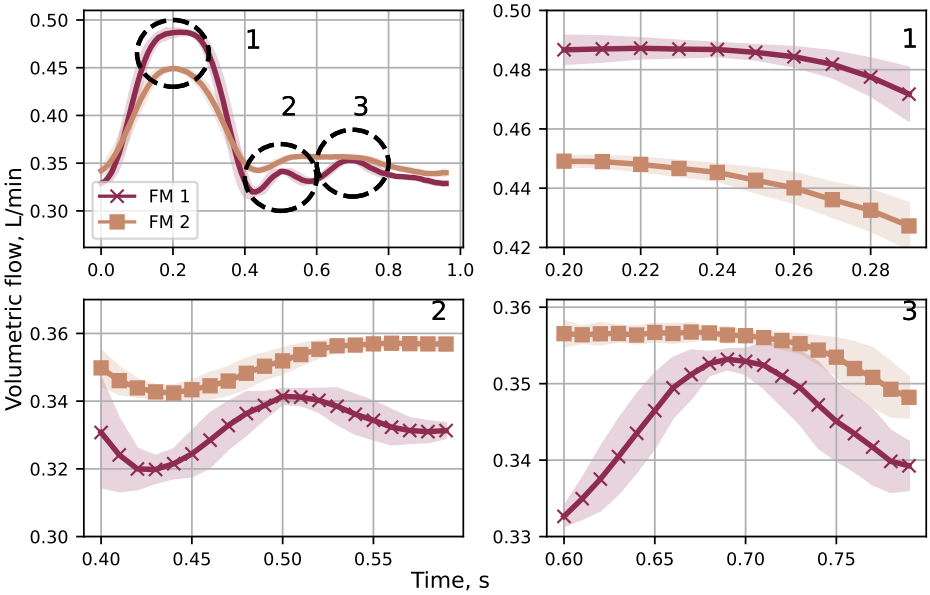

Figure 4.7: Time- and measurement-averaged representative flow curves for both flow meters with 95% confidence intervals (top left figure) along with zoomed-in regions, each indicated by a number corresponding to the region in the overall plot. (reproduced from [141])

The experimental procedure was repeated for various experimental specimens and conditions, and data was collected for many cycles in a given experiment. A visualization of the data structure can be seen in Figure 4.9, clearly showing a hierarchical, nested structure. Consequently, the measured data may vary due to factors beyond equipment noise or sensor error. Cycle-to-cycle and experiment-to-experiment variations could be attributed to various uncontrollable factors during experimental set-up such as material fatigue, varied filling of the reservoir or aquarium with artificial artery, etc. To assess the impact of the aforementioned effects, a statistical analysis of the experimental data was conducted. The variability of the data between cycles and experiments was analysed with a linear mixed modelling approach to partition the variability of

the data into components due to residual error, cycle-to-cycle variations and experiment-to-experiment variations. This decomposition is illustrated conceptually in Figure 4.10. Pressure, flow or displacement cycles observed in a given configuration are assumed to have essentially identical shapes but may vary up or down from the mean pressure level. Cycles within a given experiment will tend to be shifted similarly on average, but individual cycles will vary around this shifted mean cycle. Finally, at an arbitrary time point, noise and sensor error result in residual variations in the measured data.

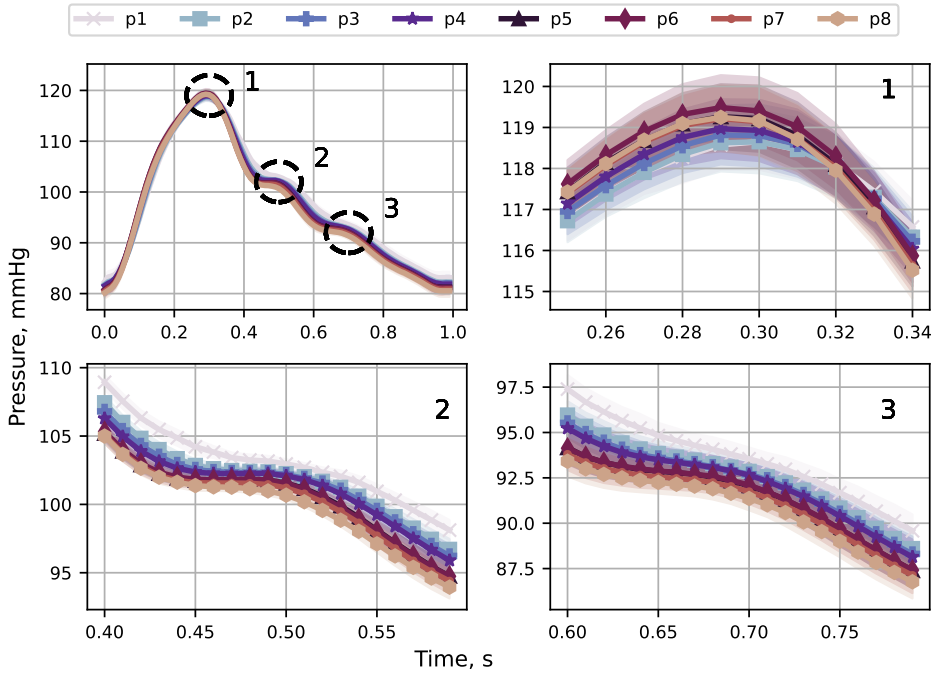

Figure 4.8: Time- and measurement-averaged representative pressure curves for all 8 transducers with a 95% confidence interval (top left) along with zoomed-in regions, each indicated by a number corresponding to the region in the overall plot. (reproduced from [141])

To evaluate the statistical significance of experiment-to-experiment and cycle-to-cycle variation, a series of models with varying complexity were compared. The analyses were conducted using the R programming language (R 4.2.0) and the lme4 package (version 1.1.30) [156]. To simplify the presentation, the focus was put on a single pressure transducer, but the overall approach

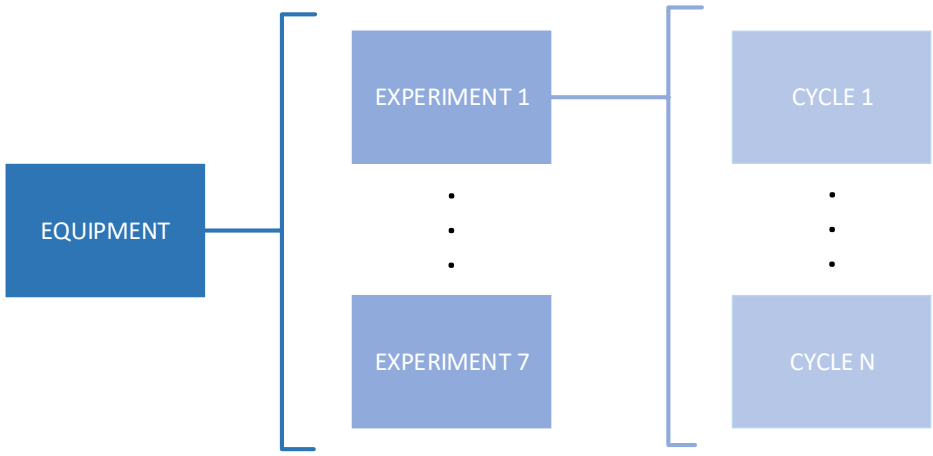

Figure 4.9: Schematic representation of data structure that resulted from experiments. For each equipment, i.e. transducer, flowmeter etc., 7 experiments were conducted, each with N measured cycles nested within. (reproduced from [141])

was applied across all pressure transducers, flow meters and cameras. Let p_{ijk} denote the measured value of pressure at position i in the k-th cycle from the j-th experiment. The residuals ϵ_{ijk} are assumed to be independently, identically normally distributed ('iid', see Section 2.4.2) with zero mean and standard deviation σ_e. First, an ordinary linear regression model (Null) is fit to the data:

$$p_{ijk} = \mu_i + \epsilon_{ijk} \tag{4.2}$$

where μ_i is the mean value of pressure at position i of the cycle. The first mixed effects model (Exp. Effect) is:

$$p_{ijk} = \mu_i + \tau_j + \epsilon_{ijk} \tag{4.3}$$

which introduces the shift τ_j for experiment j. Finally, the nested cycle effect model:

$$p_{ijk} = \mu_i + \tau_j + \alpha_{k(j)} + \epsilon_{ijk} \tag{4.4}$$

additionally introduces the shift $\alpha_{k(j)}$ for cycle k of experiment j. Again, note that τ_j and $\alpha_{k(j)}$ are assumed to be independently, identically normally distributed (iid) with standard deviations σ_{exp} and σ_{cycle} respectively. Likelihood-

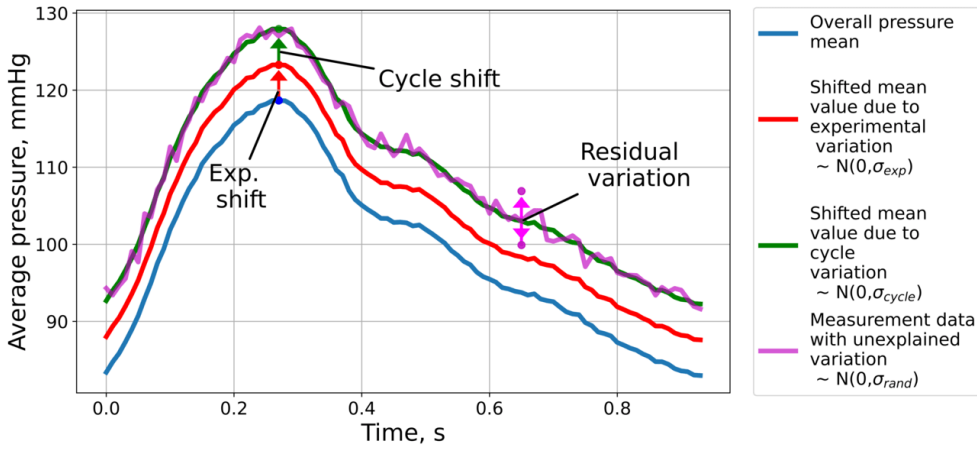

Figure 4.10: Conceptual visualization of shifts in mean value due to experimental and cycle effects as well as random unexplained variance. (reproduced from [141])

ratios [157] and information criteria [96] were used to assess, whether the variance in the data actually could be better attributed to grouping effects than residual variation (Table 4.1). Both assessments show that (4.4) better fits and explains the data than (4.2) or (4.3), and we consequently focus on the results for (4.4) in the following analysis.

Table 4.1: Reported values of model performance metrics (AIC, BIC) and likelihood-ratio tests. The bottom row of Table 4.1 shows the value in comparison to the previous model. The likelihood-ratio tests were in favor of existence of grouping effects, giving $p_{value} \approx 0$. Further, information criteria (Akaike Information Criterion, AIC, and Bayesian Information Criterion, BIC [96]) also indicate the superiority of (4.4).

Test\Model	Null Model	Exp effect	Nested cycle effect
AIC	277245.8	278164.1	253098.6
BIC	257486.4	258414	254035.7
LRT	x	$p < 0.001$	$p < 0.001$

To characterize the reproducibility of the shape of each sensor's waveform, a 95% confidence interval (CI) for μ_i was computed, while a 95% prediction interval (PI) for the mean of a new cycle, $\mu_i + \alpha_{k(j)}$, was computed to characterize

the stability of the waveform level within a given experiment [91]. The CIs were obtained using the reported standard errors of the estimated fixed effects obtained from the model fit and using the formula [158]:

$$\hat{\mu}_i \pm z_{1-\frac{\alpha}{2}} \sqrt{\mathrm{SE}(\hat{\mu}_i)} \tag{4.5}$$

for the CI of the mean cycle and

$$\hat{\mu}_i \pm z_{1-\frac{\alpha}{2}} \sqrt{\mathrm{SE}(\hat{\mu}_i) + \hat{\sigma}_{cyc}} \tag{4.6}$$

for the PI for a new cycle mean. SE denotes the standard error of an estimated parameter, and $z_{1-\frac{\alpha}{2}}$ is the $100 \times (1-\frac{\alpha}{2})^{th}$ quantile of the Standard Normal Distribution (here $\alpha = 2.5\%$ to get 95% confidence [91]). The maximum CI and PI widths, and the estimates $\hat{\sigma}_{exp}$, $\hat{\sigma}_{cyc}$ and $\hat{\sigma}_{resid}$ are reported for all equipment and cases in Tables 4.2, 4.3, 4.4 and 4.5 . The CI and PI for the full waveforms are shown in Figure 4.11 for pressure transducer p1, the inlet flowmeter, and the VEO camera in case B. The widths are 0.942 mmHg and 1.476 mmHg for CI and PI respectively, while the estimated standard deviations are 0.623 mmHg, 0.282 mmHg and 1.026 mmHg for experiment, cycle and residual variation respectively. The values for the flow waveform were 2.08E-03, 2.61E-03, 9.59E-04, 3.87E-04 and 6.77E-03 L/min, and for the VEO camera 6.10E-03, 6.22E-03, 2.10E-03, 3.05E-04, 4.81E-03 mm. The magnitudes are similar across the transducers and cases with the exception of case C for which for nearly all sensors have the highest $\hat{\sigma}_{exp}$. The flow meters have higher values in case C compared to cases A and B, while $\hat{\sigma}_{exp}$ is highest for case D. In general the widths of both PIs and CIs are narrow and confirm that the experimental procedure is reproducible and consistent. The residual variation tends to be larger than the experiment-to-experiment or cycle-to-cycle variation and suggests that high frequency noise sources are more significant than systematic variations between experiments; however, even the estimated standard deviations corresponding to these noise sources are very low compared to the mean values. While the experiment-to-experiment variation is larger in Case C compared to other estimates, some variability is to be expected in estimates of σ_{exp} as there are only 7 experiments in each case.

4.5.1 Model checking

Similarly to OLS, in LMEs there also exists the need to check the assumptions of the model so that one is sure that the inferences made are well grounded. The checks were introduced in Section 2.4.3 for OLS and can be also applied here.

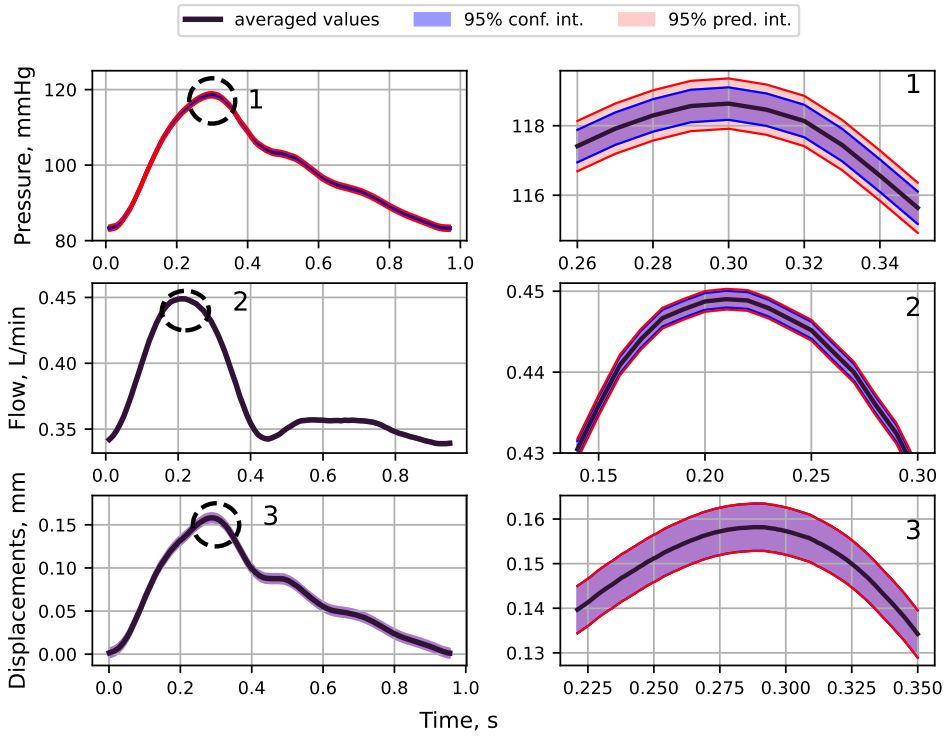

Figure 4.11: Asymptotic prediction and confidence intervals for the models fitted to pressure, flow and displacement data in Case B. The left panels show the full waveform. The regions marked by dashed circles in the left panels are shown at a higher zoom in the right panels.

First of all, the residuals of the fitted statistical model were analyzed to identify if the assumptions of independence and normality are violated. The observed residuals are defined as the difference between the predicted and measured values given in terms of (4.4) as

$$\hat{\epsilon}_{ijk} = p_{ijk} - \hat{\mu}_i - \hat{\tau}_j - \hat{\alpha}_{k(j)}. \tag{4.7}$$

A Quantile-Quantile (QQ) plot [91], distribution fits [159], residual-by-experiment vs index and a residuals-vs-cycle-time plot were used to assess the normality and independence of residuals [160]. In the QQ-plot (Figure 4.12 right) the tails of the curve stray from the Standard Normal quantiles indicating a non-normal distribution of residuals. Moreover, the fitted normal distribution has

Table 4.2: Summary of estimated cycle, experimental and residual standard deviations, along with maximum width of prediction and confidence intervals case A. Reported CI and PI widths are maximum ones (reproduced from [141])

	Case A: 110/70				
Equipment	**CI width**	**PI width**	$\hat{\sigma}_{exp}$	$\hat{\sigma}_{cyc}$	$\hat{\sigma}_{resid}$
p1, mmHg	0.622	1.530	0.409	0.349	0.953
p2, mmHg	0.625	1.514	0.411	0.344	0.962
p3, mmHg	0.643	1.515	0.421	0.342	1.089
p4, mmHg	0.649	1.511	0.424	0.340	1.108
p5, mmHg	0.652	1.528	0.430	0.345	0.860
p6, mmHg	0.671	1.546	0.442	0.347	0.912
p7, mmHg	0.658	1.527	0.433	0.344	0.938
p8, mmHg	0.648	1.520	0.426	0.343	1.007
EH1, L/min	1.817E-03	1.836E-03	7.109E-04	5.322E-04	1.147E-02
EH2, L/min	1.854E-03	1.873E-03	8.109E-04	1.144E-09	1.010E-02
MIRO, mm	2.421E-02	2.448E-02	5.608E-03	9.247E-04	2.596E-02
VEO, mm	1.631E-02	1.660E-02	5.641E-03	7.844E-04	1.829E-02

a broader peak and smaller tails compared with the estimated residuals (Figure 4.12 left). A mixture of two normal distributions ($N_1(0.027, 0.733)$ and $N_2(-0.112, 1.778)$) [161] seems to fit the residuals well, and the plot of residuals vs cycle time (Figure 4.13) reveals that the first half of the cycle tends to have more widely distributed residuals than the second half which is consistent with such a distribution. As the residuals seem to be distributed symmetrically, the statistical inferences based on the linear mixed effects model are likely unaffected [112].

Finally it is also worthwhile to have a look at Figure 4.15, which shows the residuals of one case (A) plotted one after another for all experiments. The top plot indicates that the Null model does not correctly explain some of the variation in the data and it shows up in residuals. By accounting for the cycle-to-cycle and experiment-to-experiment variation it is apparent that the remaining variation in the data now is due to residual (equipment) variance. To corroborate the interpretation of the residuals, ϵ_{ijk}, as sensor error, the distribution of the residuals was compared across equipment and cases. It is expected that the same type of sensors will produce similar distributions of residuals. The results are shown in a form of a probability density plot in

Table 4.3: Continued summary of estimated cycle, experimental and residual standard deviations, Case B) Reported CI and PI widths are maximum ones (reproduced from [141])

	Case B: 120/80				
Equipment	**CI width**	**PI width**	$\hat{\sigma}_{exp}$	$\hat{\sigma}_{cyc}$	$\hat{\sigma}_{resid}$
p1, mmHg	0.942	1.476	0.623	0.282	1.026
p2, mmHg	0.893	1.423	0.587	0.275	1.145
p3, mmHg	0.964	1.469	0.635	0.275	1.163
p4, mmHg	1.011	1.502	0.667	0.276	1.155
p5, mmHg	0.984	1.500	0.652	0.281	0.951
p6, mmHg	1.014	1.526	0.672	0.283	0.937
p7, mmHg	0.971	1.493	0.643	0.281	0.970
p8, mmHg	0.940	1.472	0.621	0.281	1.008
EH1, L/min	2.079E-03	2.609E-03	9.591E-04	3.874E-04	6.774E-03
EH2, L/min	2.079E-03	2.609E-03	9.591E-04	3.874E-04	6.774E-03
MIRO, mm	5.113E-03	5.316E-03	1.497E-03	3.682E-04	5.305E-03
VEO, mm	6.098E-03	6.217E-03	2.975E-03	3.048E-04	4.812E-03

Figure 4.14 and match this expectation, as the residuals are mostly clustered around 0 and their shape is consistent across cases.

An example of simultaneous displacement values from both cameras are presented in Figure 4.16 with the uncertainties estimated by the mixed model analysis. Displacement values from both cameras are generally close to each other and mostly lie either within the range corresponding to the calculated uncertainties, which demonstrates that the results are very similar across cameras and that the phantom deforms similarly along its principal axes. Given that, ultimately the goal is to apply this methodology to human subjects, it was important to compare the camera-measured displacements, to those registered by some readily available and inexpensive medical imaging equipment. Basing on nearly identical results from the two cameras, the top camera was dismantled and an ultrasound probe was installed instead. A few of the experiments were reproduced and, when overlaying one of them with the calculated uncertainties (along with uncertainties provided for the ultrasound equipment), it produces the results in Figure 4.17. It shows the displacements measured by the ultrasound (measurements conducted and data provided by Jan Juszczyk, PhD) and by the side camera, VEO, for case B. It can be seen that the mag-

Table 4.4: Continued summary of estimated cycle, experimental and residual standard deviations, Case C) Reported CI and PI widths are maximum ones (reproduced from [141])

	Case C: 130/90				
Equipment	**CI width**	**PI width**	$\hat{\sigma}_{exp}$	$\hat{\sigma}_{cyc}$	$\hat{\sigma}_{resid}$
p1, mmHg	4.577	4.826	3.056	0.346	1.373
p2, mmHg	4.562	4.810	3.046	0.345	1.443
p3, mmHg	4.583	4.828	3.060	0.342	1.506
p4, mmHg	4.581	4.825	3.058	0.342	1.493
p5, mmHg	4.569	4.822	3.051	0.349	1.301
p6, mmHg	4.593	4.846	3.067	0.351	1.290
p7, mmHg	4.579	4.832	3.057	0.349	1.303
p8, mmHg	4.566	4.816	3.049	0.347	1.305
EH1, L/min	1.178E-02	1.190E-02	7.089E-03	0.000E+00	1.356E-02
EH2, L/min	1.176E-02	1.188E-02	7.152E-03	0.000E+00	1.061E-02
MIRO, mm	2.448E-02	2.508E-02	1.220E-02	1.377E-03	2.012E-02
VEO, mm	2.123E-02	2.194E-02	1.059E-02	1.399E-03	1.777E-02

nitude of the displacement is almost the same for both devices. Curves seem to overlap very well. Moreover, the measured US values lie comfortably within the calculated camera's uncertainty, indicating that a US can be used to detect the displacements of a material with high accuracy.

This is further corroborated by numerical data in Table 4.6 which reports the average difference of point-wise, $\bar{\Delta}_{PW}$, cycle-average $\bar{\Delta}_{CW}$, diastolic, $\bar{\Delta}_{dias}$, and systolic, $\bar{\Delta}_{sys}$, displacements between the cameras (VEO vs MIRO) or between the camera and ultrasound (VEO vs US). The difference is computed as $\Delta_{[\,]} = d_{[\,],\mathrm{VEO}} - d_{[\,],Y}$ where Y is one of MIRO or US and $d_{[\,]}$ denotes the displacement compared. For $\bar{\Delta}_{CW}$, $\bar{\Delta}_{dias}$ and $\bar{\Delta}_{sys}$ the displacements are computed for each cycle then the differences between synchronized cycles are averaged. The averages include all measurements across all pressure levels. In the table, one can see that the differences are small, on the order of 10% of the measured values. The small magnitude of the differences is also apparent in Figures 4.16 and 4.17 where the measurements are generally quite close. The uncertainty arising from the ultrasound system was calculated and provided by Jan Juszczyk, PhD and is described in more detail in [141]. On average, both MIRO and US tend to produce larger displacement values than VEO,

Table 4.5: Continued summary of estimated cycle, experimental and residual standard deviations, Case D) Reported CI and PI widths are maximum ones (reproduced from [141])

	Case D: 140/100				
Equipment	**CI width**	**PI width**	$\hat{\sigma}_{exp}$	$\hat{\sigma}_{cyc}$	$\hat{\sigma}_{resid}$
p1, mmHg	0.497	1.557	0.343	0.369	1.056
p2, mmHg	0.481	1.557	0.335	0.370	0.889
p3, mmHg	0.471	1.556	0.327	0.370	0.943
p4, mmHg	0.475	1.534	0.326	0.364	1.137
p5, mmHg	0.459	1.518	0.316	0.361	1.042
p6, mmHg	0.471	1.542	0.326	0.367	1.018
p7, mmHg	0.471	1.543	0.325	0.367	0.999
p8, mmHg	0.493	1.550	0.341	0.367	1.031
EH1, L/min	3.196E-02	3.228E-02	2.123E-02	2.309E-03	2.053E-02
EH2, L/min	3.655E-02	3.825E-02	2.255E-02	2.502E-03	1.564E-02
MIRO, mm	1.119E-02	1.165E-02	2.064E-03	8.210E-04	1.409E-02
VEO, mm	1.069E-02	1.110E-02	6.064E-04	7.566E-04	1.395E-02

thus it seems that the phantom deforms slightly more in the horizontal plane than in the vertical plane. To summarize, it has been proven that the camera measurements of the phantom are reproducible, consistent between cameras and with the ultrasound. This makes it possible to employ the ultrasound to diagnostic arterial displacement detection without a significant loss of accuracy.

Table 4.6: Numerical summary of averaged differences calculated between cameras (VEO vs MIRO) or camera and ultrasound (VEO vs US). The first column, $\bar{\Delta}_{PW}$, results from a point-wise comparison, the second, $\bar{\Delta}_{CW}$, results from comparison of cycle-average displacements, and the final two columns result from comparison of the diastolic, $\bar{\Delta}_{dias}$, and systolic, $\bar{\Delta}_{sys}$, displacements. (reproduced from [141])

Comparison\Type	$\bar{\Delta}_{PW}$**, mm**	$\bar{\Delta}_{CW}$**, mm**	$\bar{\Delta}_{dias}$**, mm**	$\bar{\Delta}_{sys}$**, mm**
VEO vs US	-0.0188	-0.0139	-0.0067	-0.0124
VEO vs MIRO	-0.0113	-0.0010	-0.0018	-0.0219

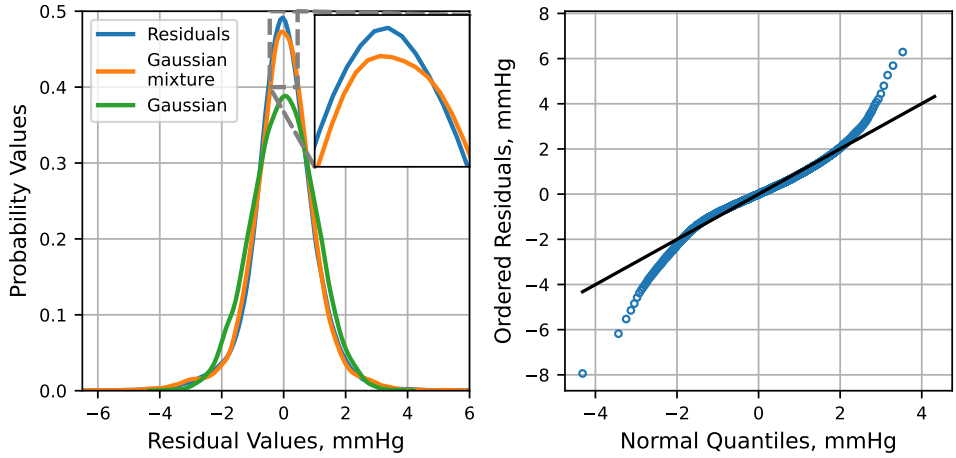

Figure 4.12: Fitted residual distributions are shown on the left with an inset showing the details of the peak. A QQ-Plot for the nested model residuals is given on the right, indicating in the tails that the residuals stray very far from theoretical quantiles. Gaussian mixture fit captures the overall distribution pretty well, except at the very peak (zoomed). (reproduced from [141])

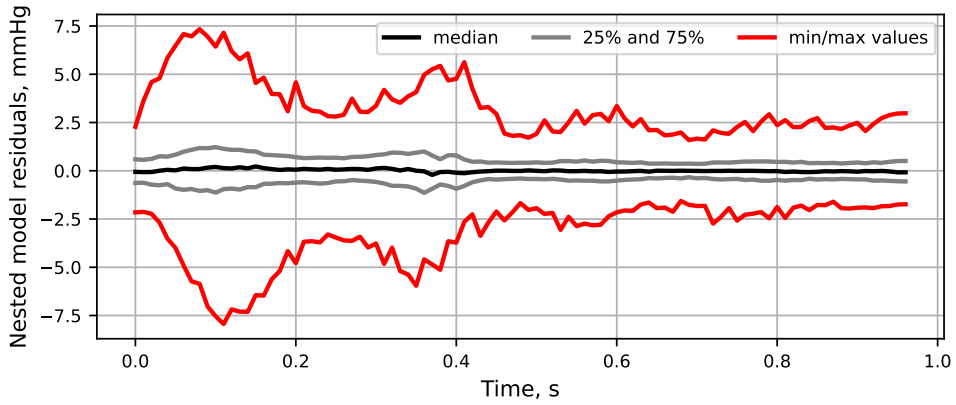

Figure 4.13: Selected quantiles of residuals of pressure for a nested model plotted versus time. (reproduced from [141])

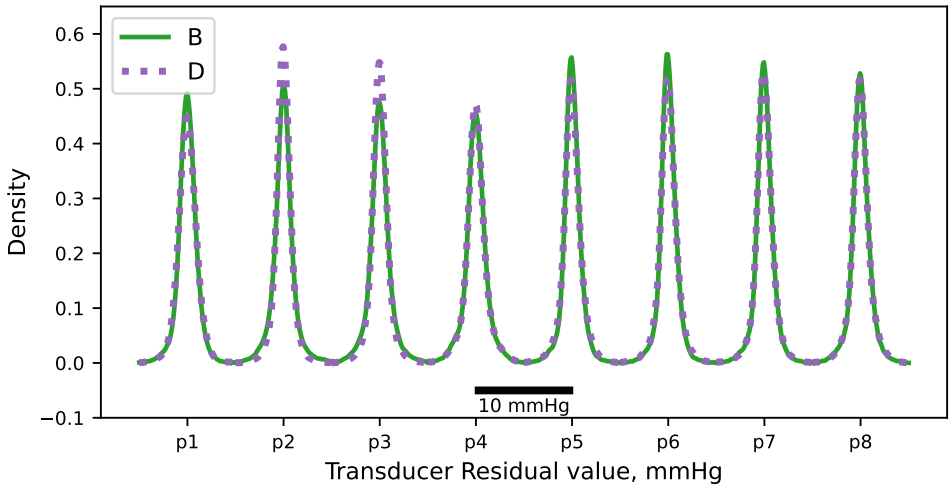

Figure 4.14: Plot of the distribution of residuals across all pressure transducers for two cases (Case B: 120/80 and Case D: 140/100). (reproduced from [141])

4.6 Analysis of Material Testing Data

To obtain data needed to verify and validate the numerical models against experimental data, there was a need to perform material testing experiments for the materials used in the experiment. Two material parameters were considered of primary interest: Young's Modulus and Poisson's ratio. Material testing experiments were performed by two independent researchers (prof. Grzegorz Kokot and prof. Wojciech Wolański). Each experiment contained multiple uniaxial tensile tests resulting in strain-stress data for tested specimen as well as transverse recordings during these tests. For each of the specimens, strain-stress curves were used to obtain Young's moduli from experimental data, while the transverse recordings were used to estimate Poisson's Ratio. To each of the strain-stress curves an intercept and slope was fitted through Ordinary Least Squares. The fit was restricted only to the linear region of the strain-stress curve. The slope of the curve was then interpreted as Young's modulus. The results from the experiments are summarized in Figure 4.18. Left side of the Figure depicts four experiments performed with provided materials, showing that the pooled mean for the Young's modulus is around 1.36 MPa . Poissons' ratios were measured only for 3 of 4 conducted experiments, they are depicted

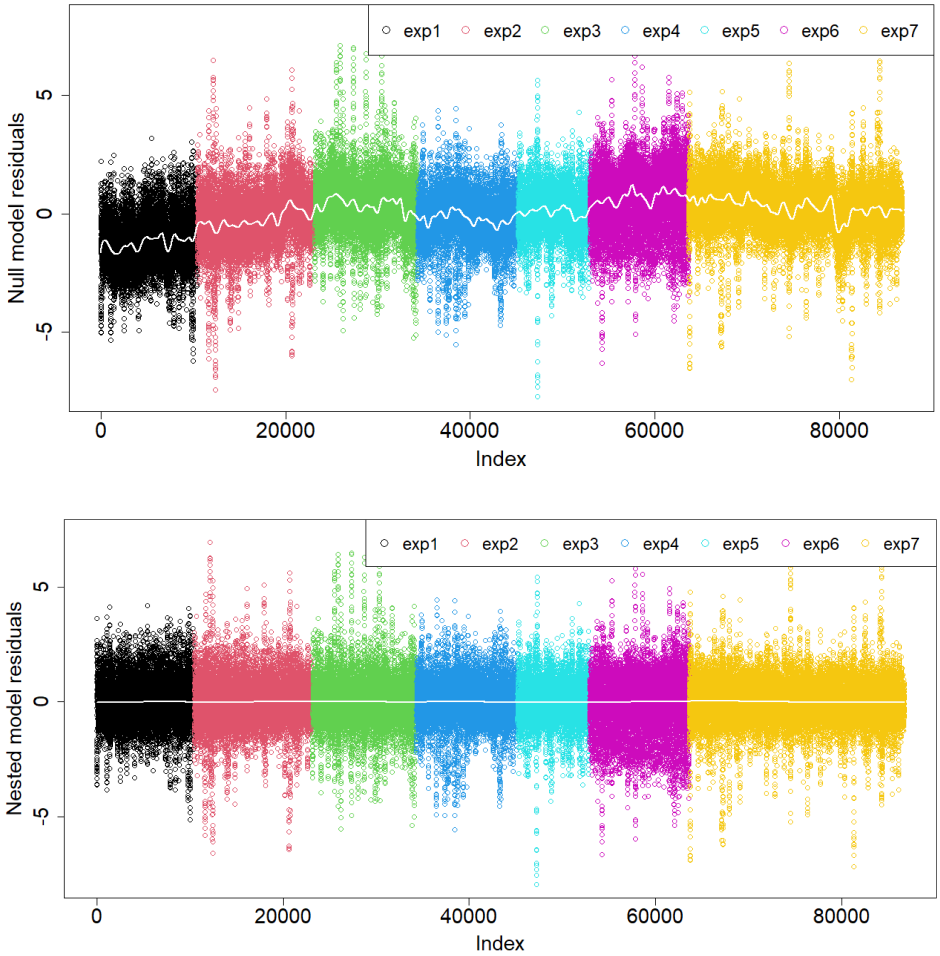

Figure 4.15: Residual vs index plot where different colors show different considered experiments within a case. **Top**: One can see that a regular Null model does not capture all of the variation present in the problem since the residuals are skewed here and there. **Bottom**: The nested model seems to capture all of the unresolved variance previously due to the fact that now the residuals for all experiments are neatly oscillating around the 0. White line in both images is the median.

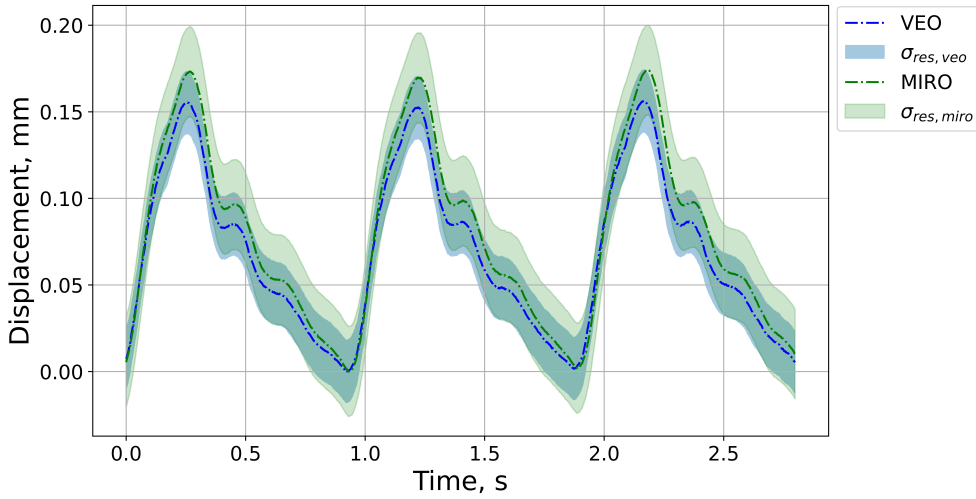

Figure 4.16: Comparison of displacements recorded by both cameras with uncertainty represented by overlaid shaded areas. In both cases the uncertainty is characterized as $d_{rel,cam} \pm \hat{\sigma}_{res,cam}$ (reproduced from [141])

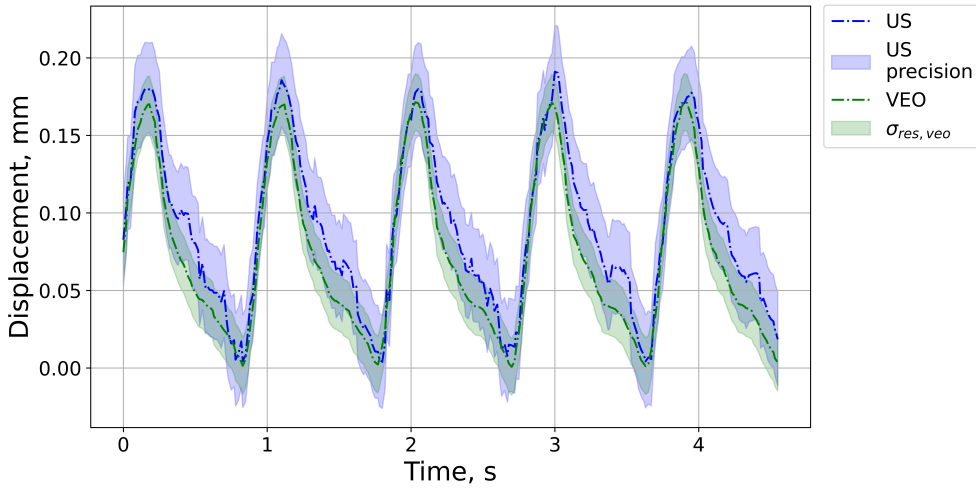

Figure 4.17: Comparison of displacements recorded by the US and VEO camera. VEO's uncertainty is characterized as $d_{rel,veo} \pm \hat{\sigma}_{resid,veo}$, while US system's uncertainty is characterized $d_{rel,us} \pm precision$. (reproduced from [141])

on the right hand side in the Figure 4.18. Similarly a pooled mean estimate was calculated and turned out to be 0.385.

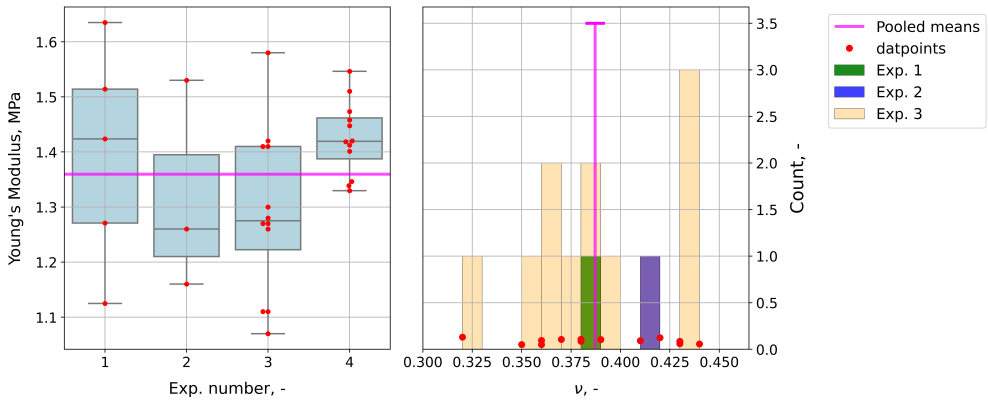

Figure 4.18: Experimental data for Young's modulus and Poisson's ratio of the arterial phantom. Left side shows a boxplot of measured Young's moduli across 4 performed experiments. Right side shows a histogram of the measured Poisson's Ratios (due to lower experiment count, a boxplot wasn't used) Jitter was applied to make points more easily distinguishable.

4.7 Problem set-up using ANSYS and mesh sensitivity

The meshed fluid domain for the phantom case can be seen in Fig 4.20. The geometry is a simple cylinder and consists of an inlet, outlet and a wall. The fluid domain is modeled using ANSYS Fluent. For the wall a no-slip boundary condition is applied. For inlet, a time-varying velocity parabolic flow profile boundary condition is applied. For outlet, a similar time-varying pressure boundary condition is applied. Both of the boundary conditions that are applied, they are reproducing the averaged flow and pressure data registered by the testing rig, described in Section 4.1. Turbulence is modeled using the Reynolds Averaged Navier-Stokes framework with the model for closure equations being k-omega SST. Due to the fact that Two-Way FSI is considered, a dynamic mesh needs to be utilized. First a coupling region was established at the domain's walls to enable data transfers between solvers. The dynamic mesh specified for the entire fluid domain uses a smoothing approach implementing a Spring/Laplace option, which treats the deforming domain as if it consisted of a series of springs

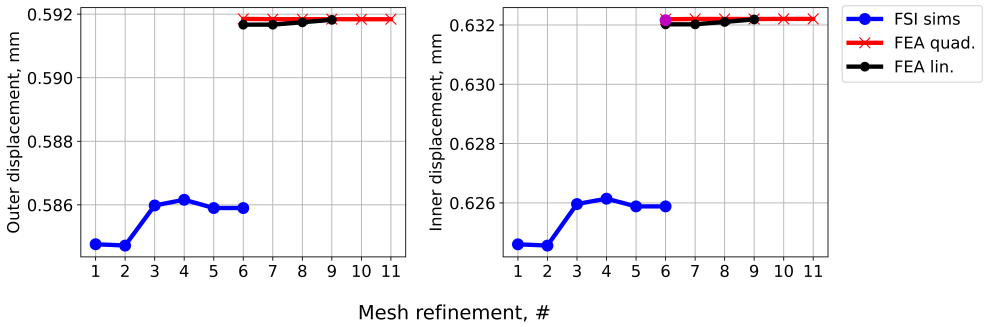

Figure 4.19: ANSYS mesh sensitivity study results. Left plot shows the outer displacement, the right plot shows inner displacements, both in mm. Data is sampled at midpoint at the timestep that produced the biggest difference in results. Blue curve shows the FSI sensitivity, red and black ones show FEA only sensitivity.

connected together. The problem naturally is a time-dependent one and uses a coupled pressure-based solver. Spatial discretization is Least Squares Cell Bases for gradients. Second Order for pressure, Second Order Upwind for Momentum, Turbulent Kinetic Energy (k) and Specific Dissipation Rate (omega). Transient formulation is Second Order Implicit. The model employs a standard initialization with zero values for pressure and x-y-z velocities. Initial values for k and omega are calculated by the software upon specifying boundary conditions.

After confirming that the problem was symmetric in its results, one location (the midpoint of the domain) was chosen for mesh sensitivity comparisons. The parameters that were varied during the study was the amount of mesh elements on various edges; edges contributing to the leading axial, radial and azimuthal directions of the tube domain. [13] Mesh sensitivity study for ANSYS simulations was performed in two parts. First 5 simulations were run. After observing that the results were flattening out, a denser fluid mesh was used while keeping the solid mesh constant. This did not change the predictions for outer displacement, so in order to save wait times and computational resources, only the solid mesh was tweaked onward. Upon performing simulation on mesh refinement #6 and observing that the values remained steady, a batch of FEA only simulations with finer meshes was run. This resulted in steady, but different result. It was noted that the difference between the finest FSI mesh and least fine FEA mesh

for the time step producing the largest discrepancies was on the order of 0.006 mm ($\sim$ 0.96%) for inner displacement and $\sim$ 0.0016 mm ($\sim$ 0.29%) for the outer displacement. The differences are lower than the resolution of ultrasound and cameras used in the experiments [141], thus they were deemed acceptable for validation purposes. These results can be seen in Figure 4.19. On the left, one can see the outer displacement values and on the right side the inner displacements. The blue curve corresponds to the FSI simulations, showing that the curve starts to flatten after the 3rd mesh refinement. The red and black curves correspond to checks done with FEA mesh only refinements for linear (black) and quadratic (red) element orders. Moreover the a relevant summary of details regarding the FSI and FEA meshes' element and node counts, as well as relevant quality metrics, may be seen in Tables 4.7 and 4.8. From the tables it can be seen that even though the quality metrics were changing (and mostly improving) with each mesh refinement, the improvements were not very dramatic. After performing the study, it was decided to stick with FSI mesh refinement #3 due to its relatively low cell and element counts but more than acceptable quality of results. This mesh refinement can be seen in Figure 4.20 (showing the fluid mesh) and in Figure 4.21 (showing the solid mesh). Both meshes use the same axial and radial spacing - this is done in order to minimize the need for interpolation of values between non-conformal meshes in the contact region. [13] As can be seen in Figure 4.20, the fluid mesh employs the O-grid design in order to maximize its quality [162].

Table 4.7: Summary of both solid and fluid meshes used for partitioned FSI in ANSYS. This table shows element and node counts as well as orthogonal qualities for both solid and fluid meshes.

ANSYS FSI Meshes	**Element count**		**Node count**		**Ortho quality (avg)**		**Ortho quality (sd)**	
Mesh no.	**Fluid**	**Solid**	**Fluid**	**Solid**	**Fluid**	**Solid**	**Fluid**	**Solid**
Mesh #1	117150	13200	121656	73080	0.983	x	4.18E-02	x
Mesh #2	156555	17640	162208	97500	0.983	x	4.18E-02	x
Mesh #3	223440	23520	231028	130000	0.985	x	4.22E-02	x
Mesh #4	294000	47040	302068	224560	0.985	x	3.90E-02	x
Mesh #5	440000	52800	451061	264880	0.985	x	3.90E-02	x
Mesh #6	440000	70400	451061	335520	0.985	x	3.90E-02	x

Table 4.8: Continuation of the table summary of both solid and fluid meshes used for partitioned FSI in ANSYS. This table shows other relevant mesh quality metrics such as skewness and element quality for both solid and fluid meshes.

ANSYS FSI Meshes	**Skewness (avg)**		**Skewness std**		**Element Quality (avg)**		**Element quality (sd)**	
Mesh no.	**Fluid**	**Solid**	**Fluid**	**Solid**	**Fluid**	**Solid**	**Fluid**	**Solid**
Mesh #1	7.94E-02	3.33E-02	9.432E-02	0.00E+00	0.119	0.218	4.184E-02	6.252E-03
Mesh #2	7.94E-02	3.33E-02	9.432E-02	0.00E+00	0.196	0.342	6.696E-02	8.367E-03
Mesh #3	6.81E-02	2.50E-02	9.042E-02	0.00E+00	0.145	0.285	5.749E-02	6.772E-03
Mesh #4	7.20E-02	2.50E-02	8.865E-02	0.00E+00	0.114	0.196	3.818E-02	5.207E-03
Mesh #5	7.20E-02	2.50E-02	8.865E-02	0.00E+00	0.225	0.361	7.357E-02	7.017E-03
Mesh #6	7.20E-02	2.50E-02	8.865E-02	0.00E+00	0.225	0.277	7.357E-02	5.691E-03

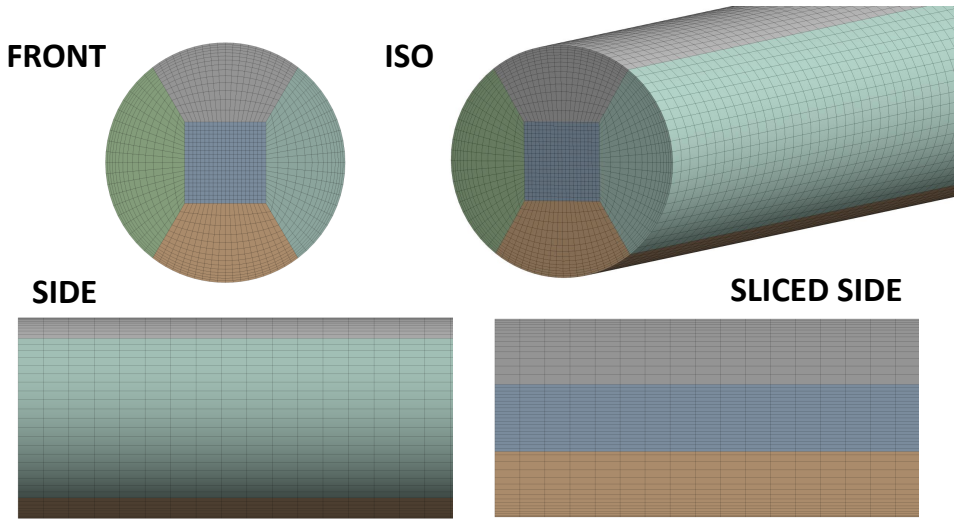

Figure 4.20: Various views for the fluid mesh (ANSYS). As can be seen an O-grid was created for the fluid domain. Four pictures showing the final mesh in various perspecitves are present (front view, isometric view, side view and a side view of the sliced mesh).

4.8 Comparison of results between different models and with phantom data

To develop a surrogate model one needs to start with experimental design. As described in the Introduction (and in Chapter 3), to achieve this it is necessary to carry out a plethora of simulations. Due to difficulty of automating the ANSYS coupling, it was necessary to rethink its use for the problem at hand. Another problem encountered while using the ANSYS software, is it became very unstable with Young's Moduli lower than 1.5 MPa and thus necessitated a lot of manual control and tuning to make it work. This instability was ascribed to the segregated approach as the results were demonstrated to be grid independent and sufficiently small time steps and large iteration counts (both within couplings and inbetween them, i.e. per time step) have been ensured to make the exchanged forces stabilize within the timestep. The alternative presented itself in the open-source package, Finite Elements for Biomechanics. [12] Given that the ANSYS package is an industry standard it was decided to compare the two approaches - FEBIO and ANSYS - against each other

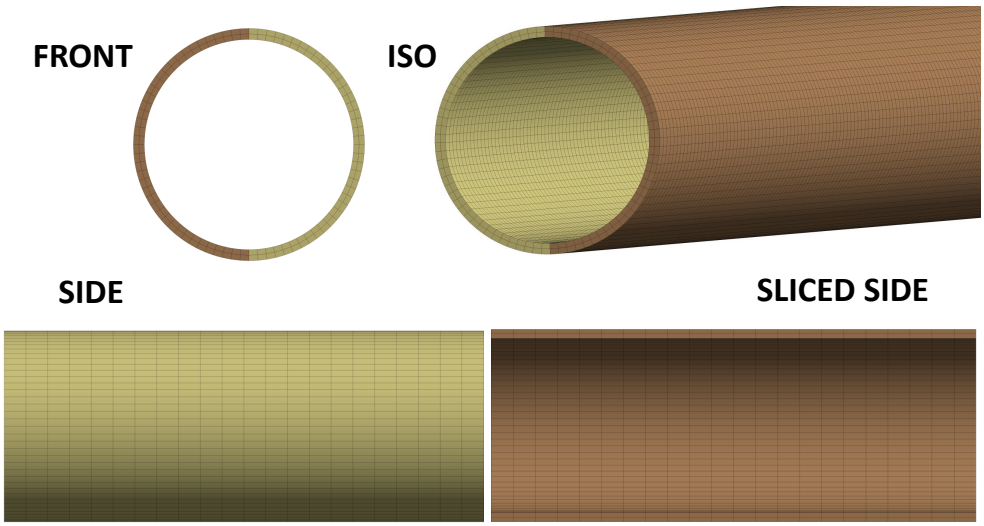

Figure 4.21: Various views for the solid mesh (ANSYS). Four pictures showing the final mesh in various perspecitves are present (front view, isometric view, side view and a side view of the sliced mesh).

and if they produced similar response, move on with just FEBIO. The results can be seen in Figures 4.23. Here, both models were evaluated at a relatively stiff and 'safe' Young's Modulus value of 1.6 MPa and Poissons ratio from Section 4.6. The image clearly shows that the Neo-Hookean model is a very close approximation of the results produced by the ANSYS model at given conditions. Isotropic Elastic seems to undepredict the displacements results, but given that this is a theoretical exercise it was decided to also test the model on actual laboratory data. The FEBIO mesh was set to reproduce the mesh #3 in the ANSYS sensitivity study.

The domain, mesh and boundary conditions used in FEBIO (especially for surrogate development) can be seen in Figure 4.22. Here, like with the ANSYS problem set-up, the domain used in this work is was a cylinder approximating the arterial phantom used in the laboratory rig. With the inner domain being the fluid and the outer domain being the solid body. Volumetric flow inlet and pressure outlet are assumed and a no-slip condition at the wall. Like in ANSYS set-up, The solid domain assumes fixed ends, i.e. no displacements at the hollow cylinder bases and the boundary conditions from inflow and outflow from the experimental measurements were used. Later, to facilitate the evaluation of

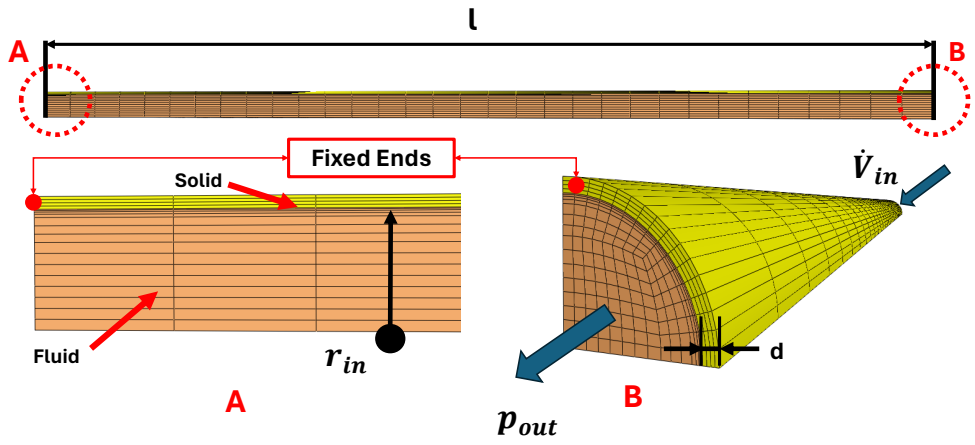

Figure 4.22: Summary image showing most relevant dimensions of the considered domain (black arrows), inlet and outlet boundary conditions (blue arrows) and other conditions and regions (red arrows) of the FEBIO model.

the model for DOE purposes, the solid and the fluid domains were reduced to a quarter of a cylinder, assuming symmetry. Both mesh and geometry were regenerated at every viable parameter combination sampled (see next Chapter, Section 5.1).

However, before proceeding with the FSI model automation it was still necessary to test whether the model reproduces the experimental data. Thus in total 4 simulations were run on the mean Young's modulus (1.36 MPa) and Poisson's ratio (0.36) estimated from experimental data. The reason for using mean values is due to the fact that the phantoms themselves were extensively strained during measurements and they were needed in their entirety, thus it was opted to measure a few other available phantoms and their scraps. These representative values were fed into the FEBIO software model definitions. Having run all four pressure ratios , 110/70 mmHg, 120/80 mmHg, 130/70 mmHg and 140/100 mmHg, for many repeated cycles until the transient effects due to simulation start and ramp up disappeared, the results can be seen in Figure 4.24. In the Figure one can see averaged displacement data measured by VEO and MIRO cameras with overlayed 95% Confidence Intervals estimated and

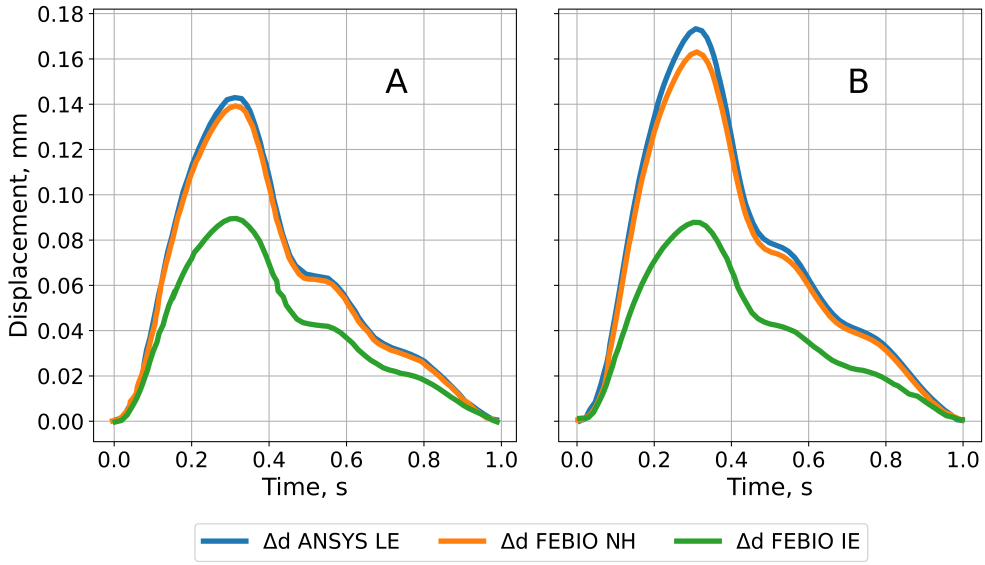

Figure 4.23: Comparison between the displacements produced by ANSYS linear elastic large deformations (blue curve), FEBIO Neo-Hookean (orange curve) and FEBIO Isotropic Elastic (green curve) for the same two experimental cases. **A** shows displacements produced for Experimental Case A: 110/70 pressure ratio and **B** shows results produced for Experimental Case D: 140/70. Plots show very good agreement between FEBIO N-H and ANSYS' LE LD model, with FEBIO IE model underpredicting the displacements.

tabulated in Tables 4.2 - 4.5. Along with them, plotted one can see the displacements produced by FEBIO's Neo-Hookean and Isotropic Elastic models. Having a look at Figures 4.24 **A**, **C** and **D** one can see that the Neo-Hookean model is slightly better than the IE model at capturing the displacements produced by the curves. In Figure 4.24 **B** the produced displacements are less than ideal for the Neo-Hookean model while being slighly better for the IE one. In the end it was decided to stick to the Neo-Hookean model and use it in the future simulations. The model delevoped here was subseuqnently used for development of the automation code of the FEBIO solver for the purposes of carrying out the Design of Experiments. The automation was achieved with an in-house developed code employing Python version 3.9.

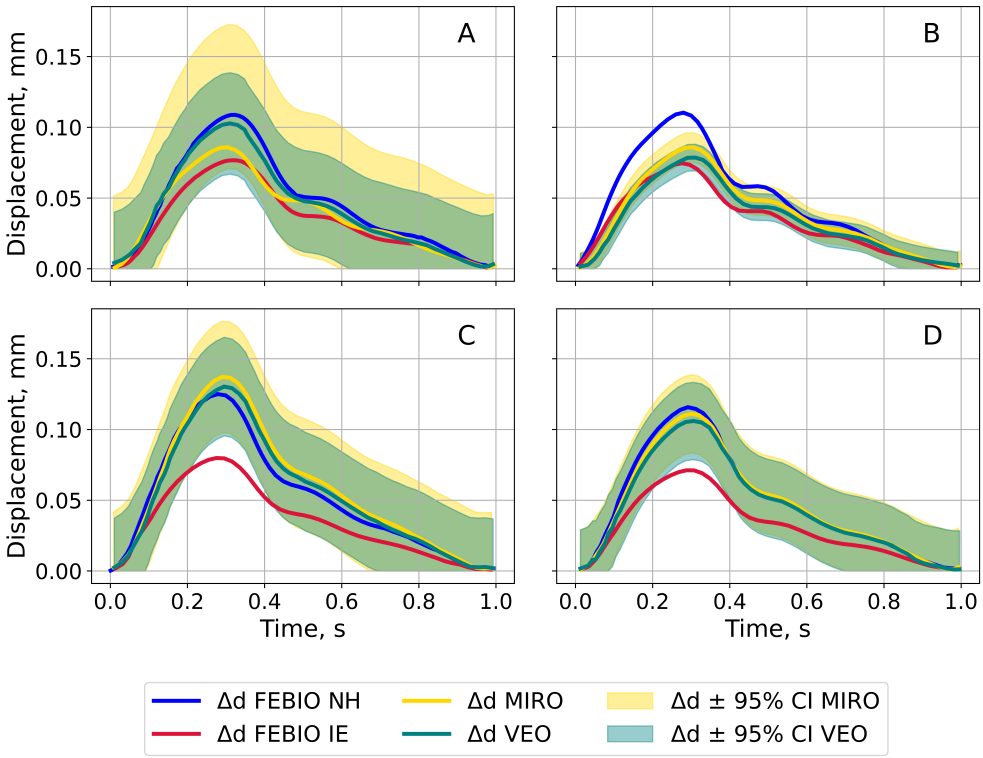

Figure 4.24: Comparison of the displacements predicted by FEBIO Neo-Hookean (blue) and Isotropic Elastic (crimson) for four experimentally tested cases: Figures **A**, **B**, **C**, and **D** correspond to their laboratory tested conditions of 110/70, 120/80, 130/90 and 140/100 respectively. The cycle- and measurement-averaged displacements captured by cameras, along with their 95% confidence intervals indicate satisfying overlap between FEBIO N-H model and Cases A,C, and D. **Note** : The color green results from overlap of experimental data's uncertainty intervals. The intervals were trimmed, such that they start at 0 displacement.

5

Surrogate development and application

As mentioned in the Introduction, tasks such as Uncertainty Quantification, Sensitivity Analysis or Parameter Estimation are very computationally demanding. The reason for that stems from their need for multiple model evaluations in order to acquire converged and reasonable estimates of the response's moments, Sobol Indices [72] or desired parameters [92, 8]. This problem can often be solved by developing a surrogate model that accelerates prediction. The development of a surrogate begins with the design of the experimental space. Once the relevant parameters and variables of interest (some correlated) are identified, their uncertainty must be assumed, either from experiments, experience or literature data. What proceeds is the sampling of the (often multidimensional) experimental space, i.e. multiple evaluations of the considered Full Order Model (here FSI) at various parameter combinations, that build up the space of solutions (so called 'response surface' [94]). With this the development of the surrogate can proceed by a proper selection of the emulator itself. This work considers two of the most recently applied approaches [123], that is Gaussian Process Regression [115] and an SVD-based Reduced Order Model. [124] It is necessary to stress that even though surrogate models enable fast prediction and facilitate the aforementioned tasks, they are not perfect. First of all they are not free of the drawbacks common for other statistical, especially

non-parametric, models. The primary drawback of both GPR and ROM developed in this Chapter, is that they do not recover the underlying phenomenon that causes the predicted response. This means that there is no actual analytical or correlation formula that is recovered and can be then simply written in a "$y = ax + b$" form, easy for interpretation and implementation. Secondly, often surrogate models require a large number of evaluations. This is due to the fact that for a model to learn the patterns in the data, it needs a lot of data and this was also the case in this problem as data set generation is generally time consuming.

After developing the models and determining the superior one, the model is employed to tasks of UQSA and Parameter Estimation. Before implementing the model within the UQSA framework, first the implementation of one of the estimators is contrasted against a common benchmarking function used in UQSA literature the Ishigami function [163]. Then it is also compared against a different method altogether - Polynomial Chaos Expansion, sort of a double check. What follows is an implementation of the surrogate model as a predictor within the Sobol Index estimators in classical Monte Carlo-based Sensitivity Analysis. This is done so that the non-influential parameters on the model output variance can be fixed and the statistical treatment can be narrowed down to only, hopefully, a handful of them. This leads to the reduction of the model complexity and the estimation of parameters of interest embedded within a Hierarchical Bayesian Parameter Estimation.

5.1 Design of the experimental space

The variables applied in the experimental design, were selected basing on physical quantities that govern a Fluid-Structure interaction problem. The boundary conditions in an FSI problem consist of (depending on the specification) volumetric flow at the domain inlet, $\dot{v}$, pressure at the outlet, P, and fluid material properties, which were kept constant in this study. Behavior of the solid domain, when considering a Neo-Hookean material, as stated above, is primarily dependent on Young's modulus and Poisson's Ratio. The deformations of the body are also governed by the inner radius of the domain and its thickness and length. [109] To better represent the variations in flow conditions, among hypothetical patients, it was decided to also make the boundary conditions related to flow, amenable to sampling. Thus a parameterization was chosen, basing on [164]. Both pressure and flow curves were parameterized by 3 variables:

the shift, the amplitude and the cycle time. Cycle time simply measures the average cycle length of both pressure and flow curves, which is defined as the time elapsed between successive diastoles within a given measurement series. While the cycle parameter is shared between both flow and pressure, the shift and amplitude ones are separate. In Figure 5.1 one can see what the shift and amplitude mean; the former is synonymous with p_{avg}, i.e. the average value within the series, and the latter refers to the difference between the maximum and minimum value. Mathematically:

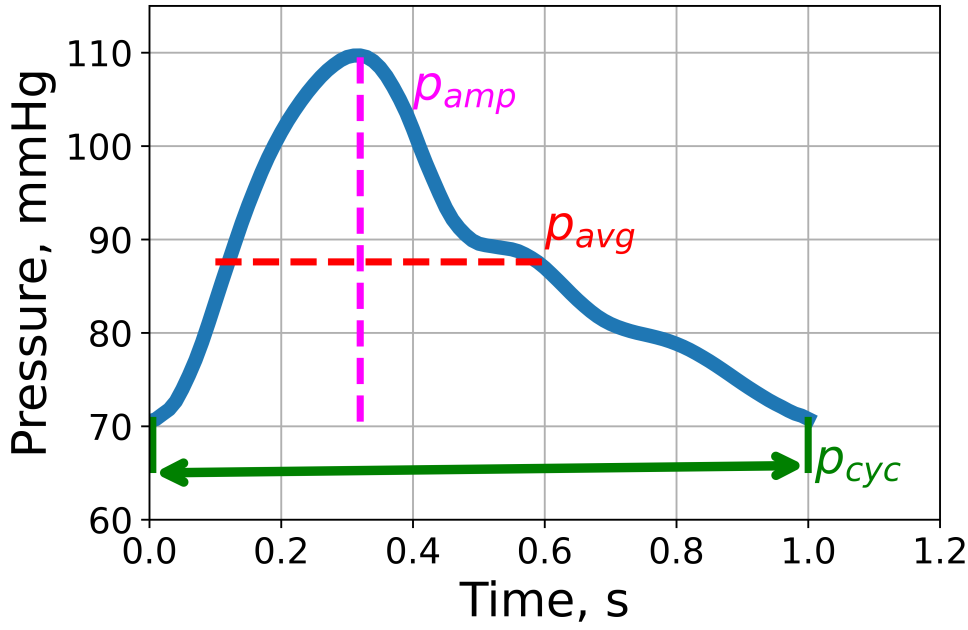

Figure 5.1: Visualization of parameterization of fluid inlet and outlet boundary conditions (here only the outlet, i.e. pressure is shown) (reproduced from [141])

$$\begin{aligned} P_{shift} &= P_{avg} = \frac{1}{N}\Sigma_i^N P_i \\ P_{amp} &= max(P) - min(P) \\ P_{cyc} &= t_{cyc} = t_{p_{dias,2}} - t_{p_{dias,1}} \end{aligned} \tag{5.1}$$

The same explanation holds for flow's shift and amplitude. While for testing

Table 5.1: Summary of the parameters and their distributions considered in the study. All parameters have units given in their most widely used format (presented in the table), they were recalculated to SI units for simulations.

Variable	Mean	SD	Source
E, MPa	1.00	0.3	exp. data,[165],
$\nu, -$	0.385	0.034	exp.data, [165]
t_{cyc}, s	0.917	0.119	[164] , exp. data
$\dot{v}_{shift}, L/min$	0.375	0.0124	[164], exp. data
$\dot{v}_{amp}, L/min$	0.249	0.0383	[164], exp. data
$P_{shift}, mmHg$	102.237	13.494	exp. data [166]
$P_{amp}, mmHg$	40.896	1.609	exp.data, [166]
r_{in}, mm	2.905	0.3013	[167]
d, mm	0.555	0.1225	exp. data [167]
L, cm	12.400	0.1250	exp. data [168]

purposes, initially all parameters were sampled independently, this naturally led to numerical instability in the solver and some unreasonable variable combinations. In order to impose structure onto sampling procedure, there was a need to induce correlations between variables that could naturally be considered to be co-dependent. Some were extracted from extensive experimental data described in Chapter 4. They were compared with scarce literature data, e.g.: the velocity and pressure shifts and the velocity and pressure amplitudes, cycle lengths, geometry length. Other parameters such as the arterial radius, r_{in}, and arterial thickness, d were calculated basing on an extensive correlation study [167]. In some cases where literature data was found and there were not insignificant differences between them, a middle ground approach was chosen by setting the sampling distribution with the median value and increasing the variance to reflect the uncertainty. Such cases pertain mostly to Young's modulus and the artery and artery phantom thickness. For example when it comes to Young's modulus, primarily in considerations for the FSI simulations, [165] used a value of around 1 MPa, while the stiffness of the arterial phantoms that were estimated from material tests in Chapter 4 proved to be larger, see Figure 4.18, and in some work values below 600 kPa [169] were reported. For correlated variables, it should be noted that such transformations were possible by employing the laws governing the multivariate distributions. Poisson's ratio was directly taken from the experiments as they gave very similar results to articles that reported it [165]. Thus to be able to test the models against real

experimental data this assumption needed to be made, nevertheless it is possible to easily extend the modeling framework to more human-like parameter values (e.g. present in [59]) if needed. To be fully explicit, the parameters may be presented in the distributional form in which they were directly fed to a sampler (units are the same as in Table 5.1).

$$\begin{bmatrix} \dot{V}_{shift} \\ P_{shift} \end{bmatrix} \sim MVN(\begin{bmatrix} 0.375 \\ 102.237 \end{bmatrix}, \begin{bmatrix} 1 & -0.409 \\ -0.409 & 1 \end{bmatrix})$$

$$\begin{bmatrix} \dot{V}_{amp} \\ P_{amp} \end{bmatrix} \sim MVN(\begin{bmatrix} 0.249 \\ 40.896 \end{bmatrix}, \begin{bmatrix} 1 & -0.145 \\ -0.145 & 1 \end{bmatrix})$$

$$\begin{bmatrix} r_{in} \\ d \end{bmatrix} \sim MVN(\begin{bmatrix} 2.905 \\ 0.555 \end{bmatrix}, \begin{bmatrix} 1 & 0.17 \\ 0.17 & 1 \end{bmatrix}) \tag{5.2}$$

$$\begin{aligned} E &\sim N(1, 0.3) \\ \nu &\sim N(0.385, 0.034) \\ t_{cyc} &\sim N(0.917, 0.119) \\ L &\sim N(12.4, 0.012) \end{aligned}$$

In total, 460 data points were sampled and set for evaluation. The number was restricted primarily by the evaluation time of the FSI model developed in Section 4.8. The variable of interest used here is the same one as in previous Chapter (4), i.e. the arterial phantom outer wall displacement. However to make modeling more manageable here, instead of the average across all outer wall displacements, now simply the maximum displacement of the wall is considered (which is located at the geometry midpoint). Given that Hammersley sequences are space filling designs it is not uncommon for them to explore the space relatively thoroughly. Thus, even though some of the variables were designed to be correlated from the start, it was still possible for the sampler to choose such an extreme combination of parameters that led to 'unphysical' results. 'Unphysical', in the context of current work, means that the produced FSI results were highly unrealistic, either from the perspective of the human body or from personal laboratory experience with arterial phantoms (see [141] and Chapter 4). Thus, based on [170], the mean transverse arterial displacement

(of the left common carotid artery) was estimated in more than 1000 subjects as $\hat{\mu}_d = 0.35$ mm with a standard deviation of $\hat{\sigma}_d = 0.28$ mm. Assuming the normal distribution with $N(\hat{\mu}_d, \hat{\sigma}_d)$, the cut-off point for the simulation results to qualify for the training procedure was relatively wide, around $\hat{\mu}_d + 2 \cdot \hat{\sigma}_d$ (not $\pm$ since displacements are generally not negative and the solver did not produce such blatantly unphysical results). This relaxed criterion covers a bit more than 95% of potential scenarios assuming normal distribution, a premise also adopted by the authors [170]. From the generated data 14 cases needed to be removed, resulting in a total of 446 used for surrogate model construction (training and testing). They can be seen in Figure 5.2. Having a closer look the cutoff can be visualized in the Fig 5.2 A, by the red horizontal line, anything above it was removed from the train/test set as per rules described above. In Fig 5.2 B one can see the proportion of the whole dataset that was taken to be the test set, i.e. 25%. It is important to note that the test split was random and the way it is presented in Fig 5.2 B is done solely for illustration purposes. Of course, before the train-test split, the obtained FSI results needed to be postprocessed to remove the ramp-up part (i.e. ramp-up from zero flow and pressure to their sampled values), which was included for the sake of stability of the numerical solver. Then they were split into individual cycles and averaged to yield the training data for the Gaussian Process Regression. To avoid potential underflow, the displacements were rescaled from meters to millimeters, and the data were passed along the pipeline to the GPR. Finally in Fig 5.2 C one can see an exemplary cycles used for model training and testing purposes.

5.2 Construction of the Gaussian Process Surrogate

The first choice to make in Gaussian Process is the choice of the kernel. By first exploring a simpler model that very roughly approximates the full order FSI model and trying multiple kernel combinations, it became apparent that the kernel combination best capturing the data was:

$$K = k_{1,time} \cdot k_{2,time} \cdot k_{3,parameters} \tag{5.3}$$

the first two premultiplied kernels allow for highly nonlinear behaviour as time progresses, while the last one jointly accounts for parameters. The simplest and so far the best choice for kernels themselves was:

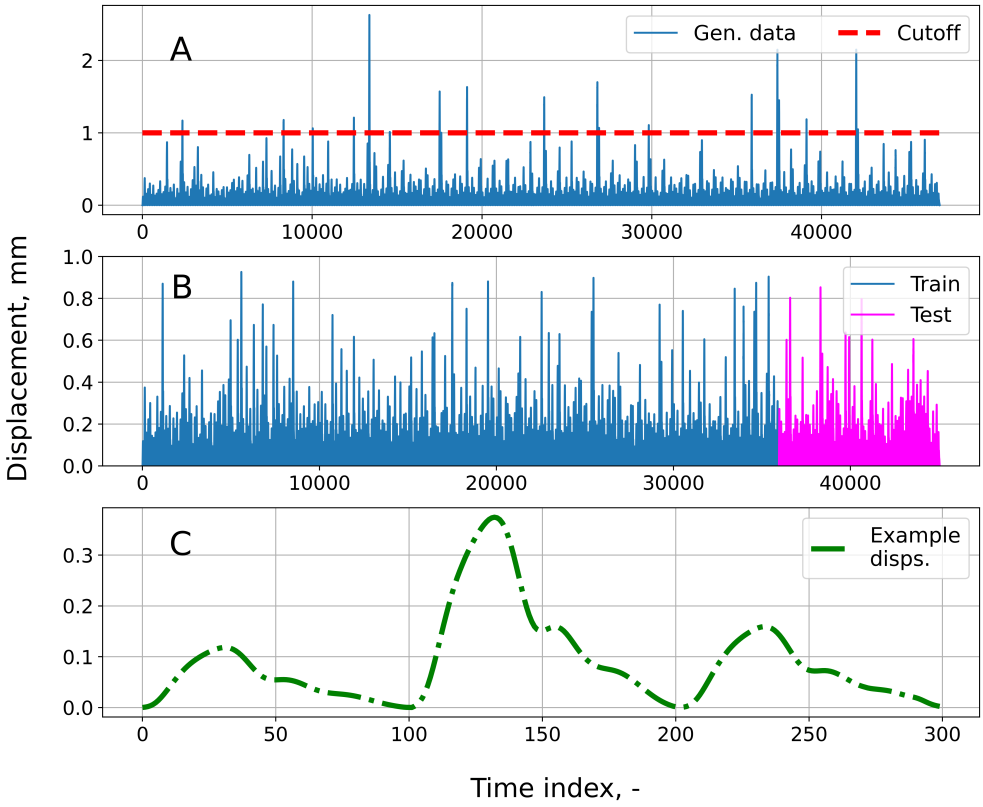

Figure 5.2: Summary of all the displacements generated from the FEBIO model through experimental design. **A** All of the results are plotted next to each other irrespective of their time ordering with a cutoff (red) that was selected to filter some outlying cases. **B** After the cutoff, displacements were split into train and test sets with the test set constituting 25% of the original dataset. **NOTE:** while the image may suggest that the data set was just arbitrarily split at some point, a random train-test set split has been performed. **C** Plot showing some example displacement curves produced. Again the X-axis reflects the order of data, not the actual time corresponding to a given displacement.

$$K = k_{RatQuad}(t) \cdot k_{RBF1}(t) \cdot k_{RBF2,ARD}(E, \nu, ...) \tag{5.4}$$

where $k_{RatQuad}(t)$ stands for a rational quadratic kernel, k_{RBF} stand for a Radial Basis Function kernel and $k_{RBF2,ARD}$ stands for Radial Basis Function

Automatic Relevance Determination (ARD) kernel. While the first two have been already described (section 3.2.1, the ARD kernel was not: its primary goal is to use one large shared kernel for many parameters with separate lengthscales for each of them:

$$k_{ARD}(x, x') = \sigma_f^2 exp(-\Sigma_{k=1}^{n} \frac{(x - x')^2}{2l_k^2}) \tag{5.5}$$

where, as previously (see Section 3.2.1) σ_f^2 is the kernel variance and l is the lengthscale. Such a formulation allows also for each of the parameters to have their own contribution to the overall model performance. Moreover it is possible to do a very 'rough', preliminary sensitivity analysis/variable importance analysis of the Gaussian Process model based solely on the inverse of the lengthscale values [171].

5.3 Sensitivity of Sparse GPs

As mentioned in Section 5.2, the key variable controlling the performance of a Sparse Gaussian Process (while controlling for training set size and kernels' hyperparameters) are its inducing variables. The sensitivity of the Sparse Gaussian Process approximation was tested with 25, 50, 100 and 200 inducing set sizes. The considered metrics were Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Median Absolute Error (MEDS) and Maximum Error (MAXERR). As mentioned in Section 3.2.2, the inducing variables are fitted alongside other hyperparameters thus in principle one could restart the optimization procedure many times and potentially obtain somewhat different estimates for them. However due to the time consuming nature of optimizing the log-likelihood in the Gaussian process, instead of bootstrapping or reinitializing the procedure hundreds of times, which was impossible due to time constraints, the error metrics were averaged along across different sample realizations and then a sampling distribution was sought for such an average. Given that the amount of training and testing samples was greater than 50, it was safe to assume that the Central Limit Theorem can be enforced and the sampling distribution would be normally distributed [67]. Thus the confidence intervals could be calculated according to the formula in Eq. 2.26, however with the T-statistic replaced by the Z-score [67]:

$$\mu \in \bar{X} \pm z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \tag{5.6}$$

The averaged metrics along with 95% Confidence Intervals can be seen in Figure 5.3. Left part of the Figure shows training set performance, while the right shows the test set. In Figure 5.3 **A** and **C** the convergence is more than satisfying for the train set (which is to be expected) with increasing sample size. The metrics seem to stabilize at around ~ 100 mark of inducing set size. The confidence intervals are very narrow (there were about 300 samples for the train size) indicating that upon repeating the fitting procedure and recalculating the metrics on the test set, most of the times the results would stay within the CIs. Looking at Figures 5.3 **B** and **C** one can see that for the test set the errors are not as small as for the train set (f.ex. at 100 mark average MAE is about three times larger for the test set than for the train set) which is normal. Also past the 100 mark it can be observed that the performance on the test set gets worse - this is a normal phenomenon and in the statistical and machine learning circles is termed overfitting. At this point the model starts to adjust too much to the training set, leading to worse generalization for unseen cases, i.e. the test set [172]. Looking at the difference in performance between 50 and 100 samples, there is not much of improvement with almost twice the inducing set size. Thus it was decided to use the model with 50 inducing variables as the go-to emulator in the following analyses. The Full Gaussian Process was not fitted for the current problem due to the fact that the size of the covariance matrix that needed to be stored in computer's RAM greatly exceeded the available resources.

5.4 Development of the Reduced Basis model

In order to obtain a reduced basis model, first one needs to have the evaluations of the Full Order Model (FOM). The considered results are the ones from Section 5.1, i.e. midpoint displacements of the arterial phantom, registered over time and for a given parameter combination. After running the FOM multiple times and removing the outliers, the results are gathered in a large matrix, often called the Snapshot Matrix and the SVD of the matrix is computed:

$$Y = \begin{pmatrix} | & | & ... & | \\ y_{\mu_1} & y_{\mu_2} & ... & y_{\mu_m} \\ | & | & ... & | \end{pmatrix} = USV^T \tag{5.7}$$

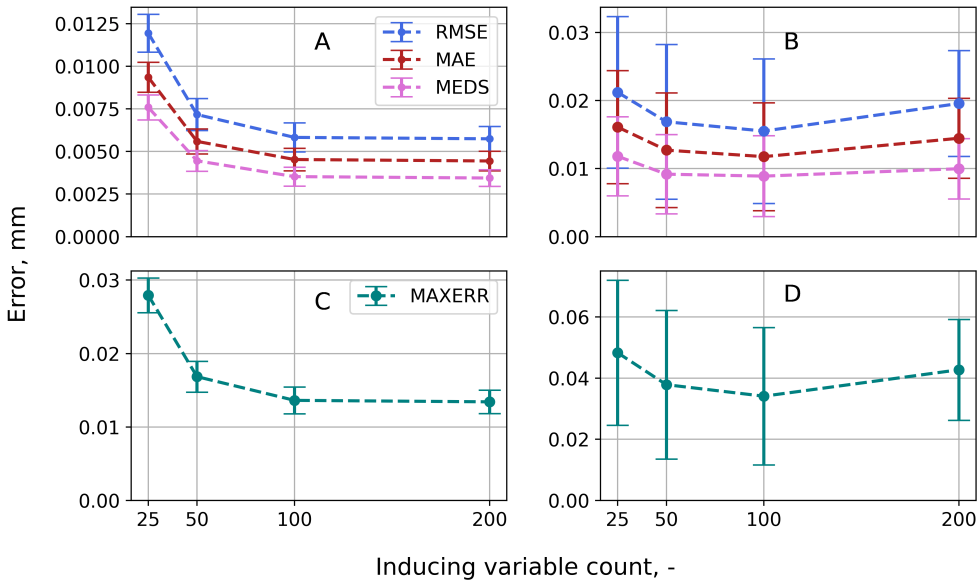

Figure 5.3: Performance of the SGPR model with increasing inducing set size. **A** and **C** show the performance of the SGPR model for the train set. It can be seen that with increasing inducing variable set size the error decreases. **B** and **D** show performance for the test set. Here it can be seen that with increasing the inducing set size at some point the model ovefits to the training data.

where y_{μ_m} is the solution vector of the $m-th$ simulation corresponding to the $m-th$ parameter combination. The amount of columns in the matrix thus corresponds to the amount of the amount of the distinct parameter combinations and the row count corresponds to time time steps. In order to obtain the reduced vector, one needs to truncate the SVD decomposition. The first few singular vectors from U can be seen in Figure 5.4 C, which in the ROM community is often referred to as an "L" curve, unsurprisingly due to its shape. The optimal choice of the singular vectors to retain can thus be eye-balled from the image, although a more systematic and quantitative approach is followed in the next section. Given that one chooses to truncate the decomposition at N, the reduced basis (RB) coefficients, q are obtained through a dot product as follows:

$$q = \hat{U}^T Y \tag{5.8}$$

where $q \in R^{r \times m}$. RB coefficients can be interpreted as a measure of similarity between the solution and the basis itself. The next step in the procedure is the development of the mapping between the input parameters and the RB coefficients. Once the data are mapped onto the coefficients, recovery of the original displacements is trivial by left multiplying equation 5.8 by U. Since the left Singular Vectors are unitary, the projection was done onto an orthogonal basis and it is not necessary to take the possible covariance of the coefficients, since there is none. This allows to fit N different full Gaussian Processes to the RB coefficients, resulting in an ensemble model.

5.5 ROM performance

Before applying the reduced basis model, it is necessary to test its performance. In order to do so bootstrapping has been employed. The same train-test sets were employed here to directly compare this model's perforamnce against the SGPR one. The used performance metrics were Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Median Absolute Error (MED). 100 bootstrap samples were used to fit 4 reduced basis models with increasing complexity (i.e. retaining the first 1,2,5 and 10 singular vectors), the metrics were then summarized across cases within one bootstrap sample using the sample quantiles. The summary of train and test set performance can be seen in Figure 5.4. In 5.4 **A** and 5.4 **B** one can see a comparison between the train and test set performance with increasing number of retained singular vectors. For the train set naturally the error diminishes with more samples, however for the test set it remains relatively the same for 100 bootstrap resamples. The explanation can be understood by looking at 5.4 **C**: here the relative contribution of each singular vector is shown. It is immediately visible that the first singular vector is the most significant one with having more than 80% of contribution (93.5% to be exact). Thus, adding subsequent singular vectors is not as informative and can even lead to overfitting and poorer model performance. In 5.4 **D** one can see a reconstruction of the displacement curve using only one singular vector. Although the reconstruction is not ideal it is still a very good one.

5.6 Model comparison

Having a look at results presented in sections 5.3 and 5.5 it is apparent that the performance of Sparse Gaussian Processes is better, on average they are 2-3

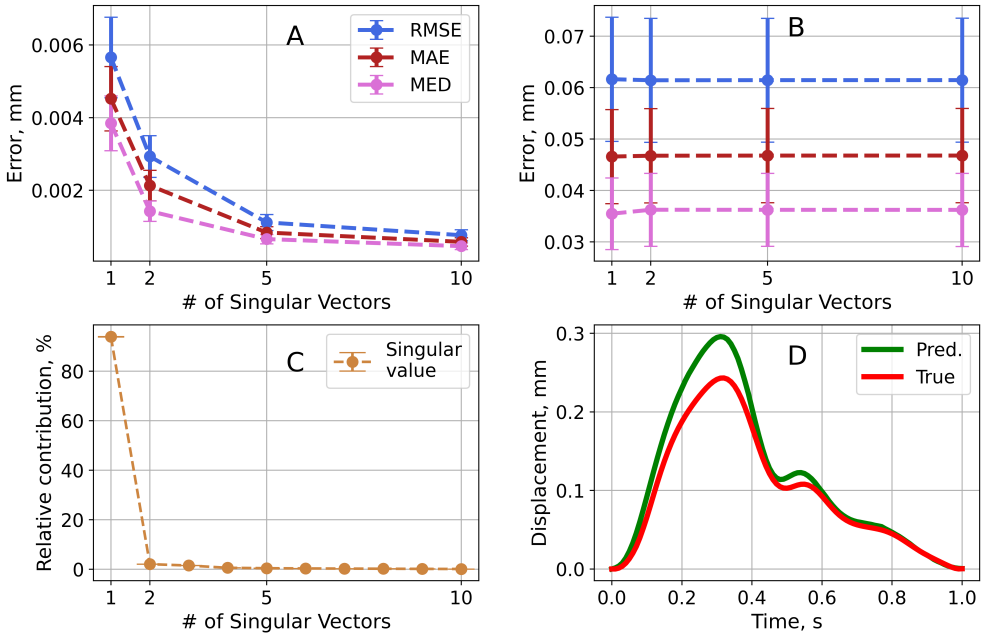

Figure 5.4: Performance of the SVD-based Reduced Order Model. **A** and **B** display the bootstrapped model performance with increasing amount of retained singular vectors. **C** shows the relative contribution of the Singular Values to the total 'energy'. **D** shows an exemplary curve reconstructed from the test set predictions.

times less erroneous. Such a state of affairs can be explained by the fact that SVD only captures linear relationships within the data [119] and the considered problem is not as simple as that: there might be some nonlinear contributions that twist and turn the displacement curves which SVD is completely missing. In further work one could potentially have a look at other methods that extend the SVD algorithm, f.ex. through kernelization techniques (e.g. kernel Principal Component Analysis) or different methods altogether, e.g. Gaussian Process Latent Variable Models (GPLVMs). One caveat is that while these methods are better at capturing the nonlinear dependencies in the data, they lack the ability to transform back and forth between the latent representation of data and the actual one. This could pose a problem when one desires to re-obtain the order reduced data in its original form. [119] Thus for the rest of the work the model developed on the basis of Sparse Gaussian Process is used.

5.7 UQSA Benchmark: The Ishigami Function

Next step was the application of the chosen Surrogate Model for performing UQSA of the Full FSI model. It is important to first benchmark the implemented methods, often using simpler and tested functions. One such common function in the SA community is the Ishigami function [163]:

$$Y = sin(X_1) + a \cdot sin^2(X_2) + b \cdot X_3^4 sin(X_1) \tag{5.9}$$

where X_i are random variables and are all uniformly distributed, i.e. $X_i \sim U(-\pi, -\pi)$, a and b are arbitrary parameters, often chosen to be $a = 7$ and $b = 0.1$ [163, 173]. The results of the benchmark can be seen in Figure 5.5. In Figure 5.5 **A** First Order Sobol indices are displyed while in Figure 5.5 **B** Total Sobol indices are shown. They are calculated for the Ishigami function, using the Sobol [74] and Homma/Saltelli [74] methods, using $N = 1000000$ samples. For completeness, they are contrasted with an entirely different method that is not based on the Monte Carlo approach, the Polynomial Chaos Expansion. To be specific the 8-th order Quadrature-based Sparse Polynomial Chaos Expansion (PCE), using $N_{nodes} = 840$ sparsely generated nodes, has been employed. Those interested in Polynomial Chaos are referred to [174]. In both cases the implemented methods approximate the literature data pretty well and reflect the actual importance of variables properly. Figure 5.5 **C** shows the histogram of the Ishigami ($N = 100000$ samples) with characteristic 'devils horns', which also is a sanity check for its proper implementation. [173] Having benchmarked the implemented methods, it is possible to move on to applying them on the actual problem at hand, i.e. the full order Fluid-Structure Interaction model, using its surrogate. It is important to note that while Polynomial Chaos Expansion is also a widely applicable method in the domains of UQSA, it is not as widely used as a surrogate model per se and thus was only chosen as a refernce point and not the main model used in the analyses. Moreover, PCE does not support the construction of model with dependent inputs as it assumes each of the random variables to be independent [174].

5.8 SA for correlated inputs

The assumption of correlated variables (e.g. v_s and p_s) in the design space definition, introduces a violation of the assumptions of input independence of the ANOVA-HDMR decomposition in Eq. 3.31. This can be handled by

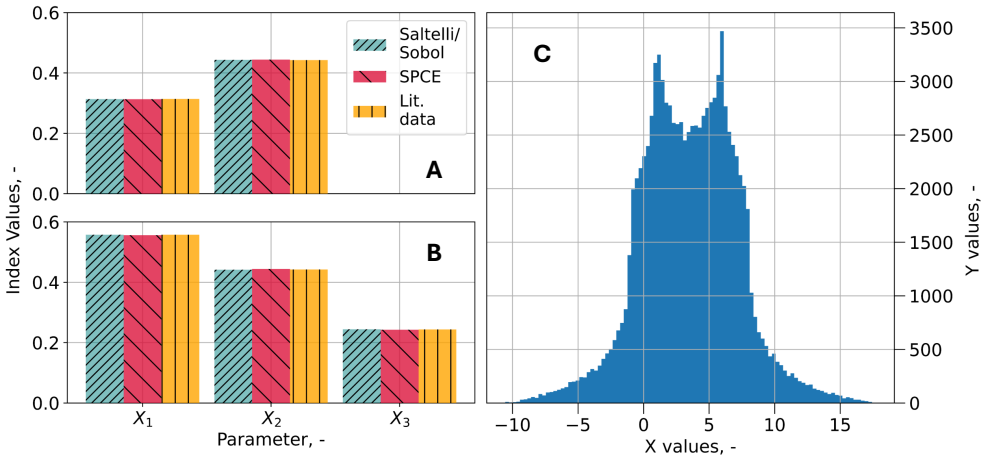

Figure 5.5: Ishigami function benchmark. **A**: First Order Sobol Indices. **B**: Total Sobol Inidices. Both calculated using one of the implemented methods, the Sobol method for First Order Indices and Saltelli/Homma method for Total Order indices. They are compared against a different method, Polynomial Chaos Expansion, and against data reported in the literature. **C**: Histogram of the Ishigami function

decorrelation (or orthogonalization) of the dependent variables. A marginal linear dependence between a (normal) set of random variables may be described by a conditional expectation, $\mathbb{E}(x_i|x_j)$. If x_i and x_j were independent $\mathbb{E}(x_i|x_j) = \mathbb{E}(x_i)$. Another way to describe the dependence is through the application of the multiplication rule [131] on the joint probability distribution of two variables $p(x_1, x_2)$:

$$p(x_1, x_2) = p(x_2|x_1)p(x_1) = p(x_1|x_2)p(x_2) \tag{5.10}$$

Given independence, the joint distribution could be factored into:

$$p(x_1, x_2) = p(x_1)p(x_2) \tag{5.11}$$

Thus, using these two definitions and setting [126]:

$$x_{2-1} = x_2 - \mathbb{E}(x_2|x_1) \tag{5.12}$$

the result is:

$$p(x_1, x_{2-1}) = p(x_1)p(x_{2-1}) \tag{5.13}$$

i.e. a decorrelation of variables and induced independence.

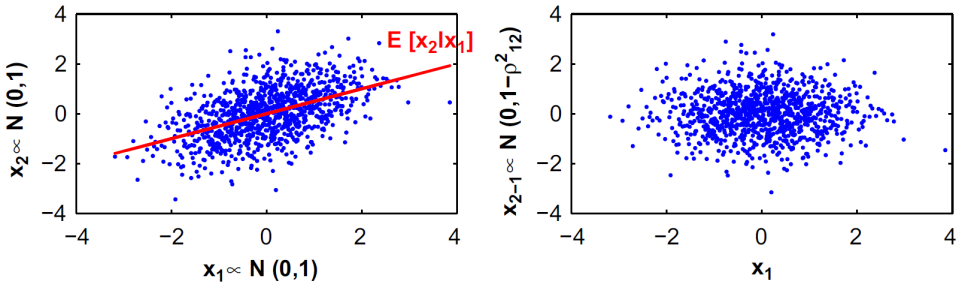

Figure 5.6: Illustration of the decorrelation procedure. By removing the linear dependence of variable x_2 on x_1 one essentially creates a new, decorrelated variable x_{2-1} on which the future inference is performed. Reproduced from [126]

The decorrelation procedure can be visualized in Figure 5.6. There, from random variable x_2 the linear dependence on random variable x_1 is subtracted. This results in the creation of the new variable, x_{2-1}, i.e. x_2 with subtracted x_1 dependence. The relation between x_{2-1} and x_1 now looks like an uncorrelated clump of points, which was the goal of the operation. Although this decorrelation makes it possible to perform SA through ANOVA-HDMR, the definitions of the Sobol indices are slightly different. In this setting, S_1 is now a full Sobol index of x_1 containing its influence in all First Order (main) effects where it is a contributing factor due to mutual dependence on other variables. $S_2 = S_{2-1}$ is the marginal index of x_2, with all of its correlations removed. [126]. Similar interpretations hold for total indices. These revised interpretations only apply to the correlated variables in the set and the old interpretation remains valid for all independent ones. [175] Using this approach, the total number of samples for the Sobol index estimates to be relatively representative is given by: [74]

$$N = n(k + 2) \tag{5.14}$$

where N is the number of evaluations required, n is some sufficiently large number ($n \geq 500$ is often a safe bet as in [72]) and k is the number of parameters. Given that in the current application the parameters are correlated this number needs to be multiplied by two, as for the correlated parameters both the Full and Independent (marginal) Sobol Indices need to be estimated. A Sobol index, e.g. S_i^{Sobol} is computed at a given time-step, i.e. $S_i^{Sobol} = S_i^{Sobol}(t)$, however for simplicity the former notation will be used. Moreover aggregated, i.e. time-averaged, Sobol indices will be presented later and denoted e.g.:

$$\bar{S}_i^{Sobol} = \frac{1}{T}\Sigma_{t=0}^{T} S_i^{Sobol}(t) \tag{5.15}$$

5.9 UQSA using the Gaussian Process surrogate

The process for UQSA utilized the fitted Gaussian Process surrogate model to enable Monte Carlo estimation of the integrals. The sampling algorithm used was Hammersley sampling. The algorithm behind the UQSA began with sampling one large, doubled parameter matrix, i.e. AB of dimension $N \times (2K)$ where K is the number of parameters. After generating the samples the matrix was split into two sub-matrices, A and B. Subsequently the samples were decorrelated using Equation 5.12, this was applied to the tuples of $(\dot{v}_{shift}, p_{shift}), (\dot{v}_{amp}, p_{amp}), (r, l)$. This split the actual Sobol indices into Full and Marginal ones as described in Sec. 5.8. Subsequently, the A_B^i and B_A^i matrices were populated. The surrogate model was then evaluated at the parameter combinations arising in the matrices A,B, A_B^i, and B_A^i. These evaluations were then fed to the Sobol index estimators in Equations 3.37 and 3.38. All these steps were then embedded into two different kinds of separate loops. As mentioned previously, the implemented methods were validated against the Ishigami function [74] and produced the same results.

The first loop tested the convergence of the Sobol index Monte Carlo estimators. In total, according to formula 5.14, around $N = 1000 \cdot (10 + 2)$, where n was chosen to be twice as usually chosen (1000 instead of 500) to be on the safe side. Thus, it was estimated that at around 12 0000 samples the convergence should be reached. The convergence was tested at 10, 50, 100, 500, 1000, 5000, 12 000, 15 000 and 20 000 samples, the convergence for two chosen parameters (E, l) was presented in Figure 5.7. It can be seen From the plots that above 5000 samples the estimates do seem to stabilize and are relatively consistent especially as one approaches 12 000 (the theoretically suggested amount), let

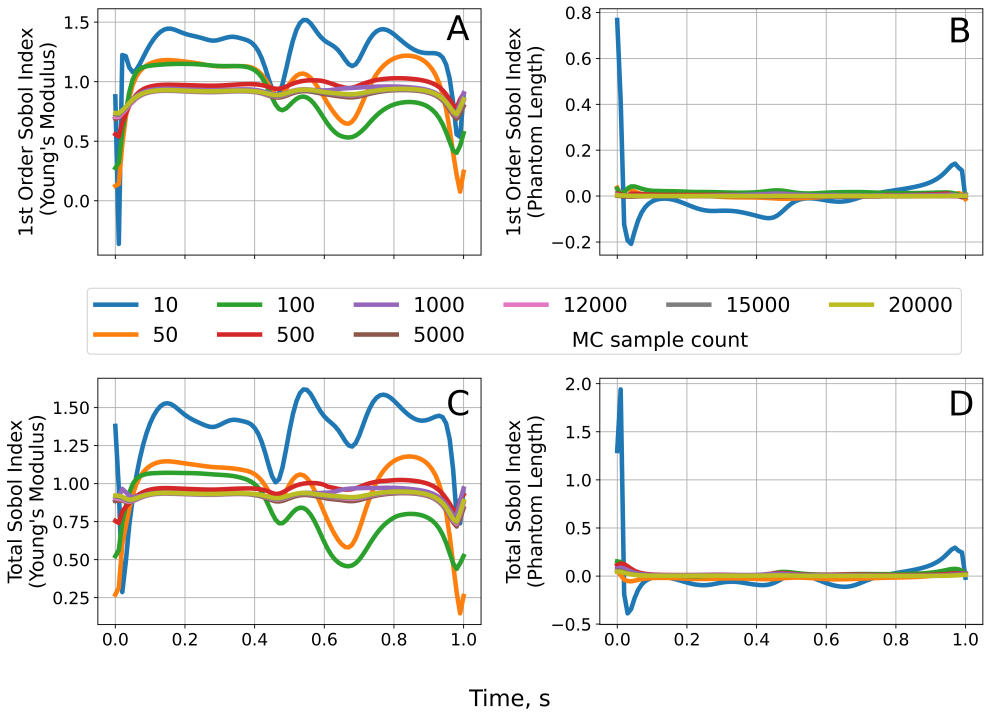

Figure 5.7: Convergence of First order and Total Sobol Indices for Young's Modulus (E) and domain Length (L). $\mathbf{A}, \mathbf{C}$ shows the convergence of First and Total Sobol Indices respectively for E. Figures $\mathbf{B}, \mathbf{D}$ show the same for domain length L. It can be seen that the convergence is easily achieved using theoretically calculated 12 000 samples, as the line corresponding to it overlaps with estimates produced from 15 000 and 20 000 samples.

alone 15 000 or 20 000. Thus with 12 000 being a relatively good sample size a different loop was employed.

The second loop relied on bootstrapping [133], i.e. resampling with replacement, the Sobol indices to get some idea of the uncertainty that they possess for the problem at hand. 100 bootstrap samples were taken in total. The median timewise Sobol indices can be seen in Figure 5.8 A,D,G for First Order Indices and B,E,H for total indices. Each row of this Figure presents results for applied methods (Sobol, Saltelli, Jansen). In Figure 5.8 C, F, I the time-averaged median Sobol indices are presented with 0.5-th and 99.5-th quantiles plotted as error bars. It can be gauged from the third column of the plot that

Young's modulus, E is the most influential parameter across all methods, which is indicated by the relative size of its Total Sobol index. Similarly, the height of its First Sobol index is dominant, indicating that this is the parameter contributing the most variance among others. The variance seems to be distributed relatively equally among the remaining parameters indicating that their overall impact on it is somewhat similar. This does not mean that these parameters should be removed - they still influence the output of the model, they are just less influential on the output's variance. While Sobol indices (both First Order and Total) by definition cannot be lower than zero, when their value is relatively close to zero it is possible to get negative values due to numerical issues, which should not be treated as a violation of the assumptions [125]. The numerical values of the estimated Sobol indices, along with their uncertainty intervals are presented in Table 5.2.

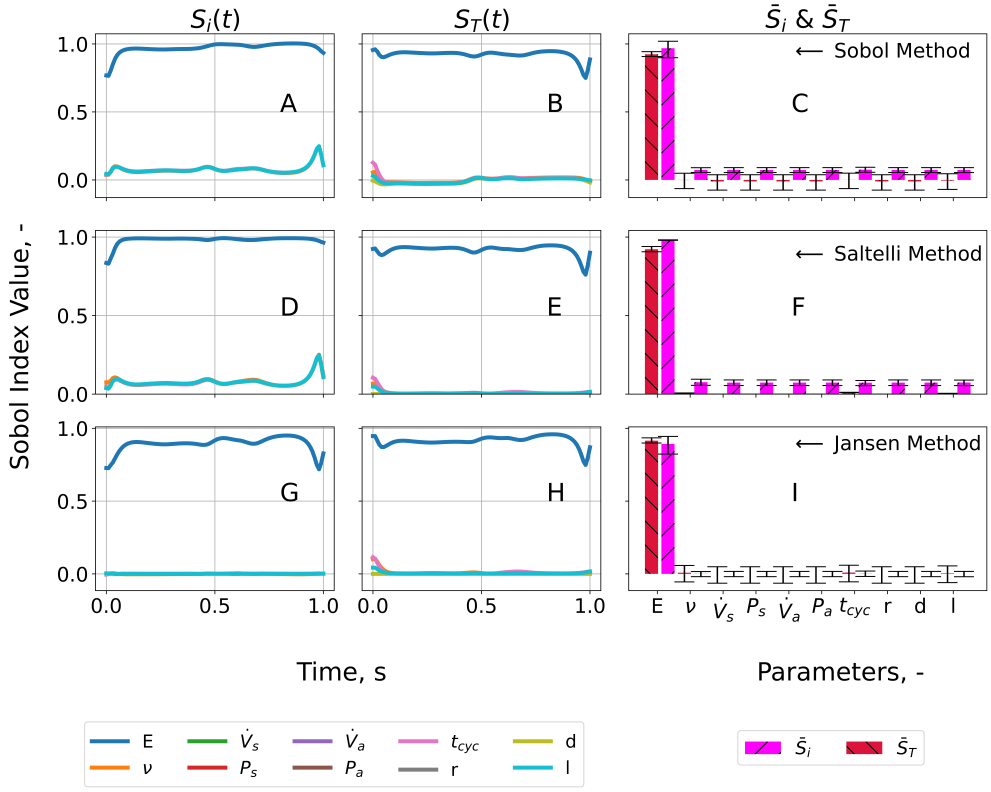

Figure 5.8: Results of the Sensitivity Analysis. Each of the rows corresponds to one of the three implemented methods for First Order/Total Sobol Indices. First row used the Sobol/Homma estimators, second row used the Saltelli/Sobol estimators and the final row used the Jansen/Jansen estimators for First Order and Total Indices respectively. Figures **A**, **D** and **G** depict the time-dependent First Order Sobol index, as can be seen it varies somewhat over time but consistently the Young's modulus emerges as the most influential parameter, meaning it should be prioritized in further research. Figures **B**,**E**,**H** show the time-dependent Total Sobol Index, here again the Young's modulus is clearly most influential, meaning that it contributes the most amount of variance to the model overall. In Figures **C**, **F** and **I** one can see the time-averaged First Order and Total Sobol Indices with overlapped 99% bootstrapped Confidence Intervals.

Table 5.2: Summary tables presenting results of the Sensitvity Analysis (intervals come from 100 sample bootstrapping). **Top**: Resulting median First Order Sobol Indices for parameters with bootstrapped 99% Confidence Intervals. **Bottom**: Resulting median Total Sobol Indices for parameters with bootstrapped 99% Confidence Intervals.

Param.	$\bar{S}_i^{Sobol}$	$\bar{S}_i^{Sobol}$ 0.15%	$\bar{S}_i^{Sobol}$ 99.85%	$\bar{S}_i^{Saltelli}$	$\bar{S}_i^{Saltelli}$ 0.15%	$\bar{S}_i^{Saltelli}$ 99.85%	$\bar{S}_i^{Jansen}$	$\bar{S}_i^{Jansen}$ 0.15%	$\bar{S}_i^{Jansen}$ 99.85%
E	0.968	0.889	1.047	0.98	0.979	0.819	0.893	0.819	0.975
ν	0.073	0.047	0.093	0.076	0.052	-0.005	-0.001	-0.005	0.003
$\dot{V}_s$	0.073	0.049	0.095	0.073	0.049	0	0	0	0
p_s	0.073	0.049	0.095	0.073	0.049	0	0	0	0
$\dot{V}_{amp}$	0.073	0.049	0.095	0.073	0.049	0	0	0	0
p_{amp}	0.073	0.049	0.095	0.073	0.049	0	0	0	0
t_{cyc}	0.075	0.05	0.098	0.073	0.049	-0.002	0.002	-0.002	0.005
r	0.073	0.049	0.095	0.073	0.049	0	0	0	0
d	0.073	0.049	0.095	0.073	0.049	0	0	0	0
l	0.073	0.049	0.095	0.074	0.05	-0.002	0	-0.002	0.003

Param.	$\bar{S}_T^{Sobol}$	$\bar{S}_T^{Sobol}$ 0.15%	$\bar{S}_T^{Sobol}$ 99.85%	$\bar{S}_T^{Saltelli}$	$\bar{S}_T^{Saltelli}$ 0.15%	$\bar{S}_T^{Saltelli}$ 99.85%	$\bar{S}_T^{Jansen}$	$\bar{S}_T^{Jansen}$ 0.15%	$\bar{S}_T^{Jansen}$ 99.85%
E	0.924	0.901	0.951	0.922	0.899	0.838	0.915	0.838	0.994
ν	-0.001	-0.083	0.077	0.007	0.006	0.002	0.008	0.002	0.014
$\dot{V}_s$	-0.012	-0.091	0.065	0	0	0	0	0	0
p_s	-0.012	-0.091	0.065	0	0	0	0	0	0
$\dot{V}_{amp}$	-0.012	-0.091	0.065	0	0	0	0	0	0
p_{amp}	-0.012	-0.091	0.065	0	0	0	0	0	0
t_{cyc}	-0.001	-0.082	0.072	0.01	0.009	0.002	0.009	0.002	0.018
r	-0.012	-0.091	0.065	0	0	0	0	0	0
d	-0.012	-0.091	0.065	0	0	0	0	0	0
l	-0.006	-0.085	0.072	0.004	0.004	-0.002	0.004	-0.002	0.01

Going back to Uncertainty Quantification itself, it was enough to gather all the intermediate system evaluations done to estimate the Sobol Indices. A collection of time-dependent responses (displacements) of the system. What was next was to calculate the time-dependent mean output of the system, along with a 99% Confidence Interval on it. The results are visualized in Figure 5.9. It clearly demonstrates that the largest uncertainty due to the parameters is present in the 0.2 - 0.4s time range which corresponds to the systole. This simulated result corroborates the experimental results from Chapter 4 where also the largest uncertainty was found in the systole region. This clearly indicates that it is worthwhile to more closely examine the systolic portion of the cycle as, on average, it is the one most laden with uncertainty [67].

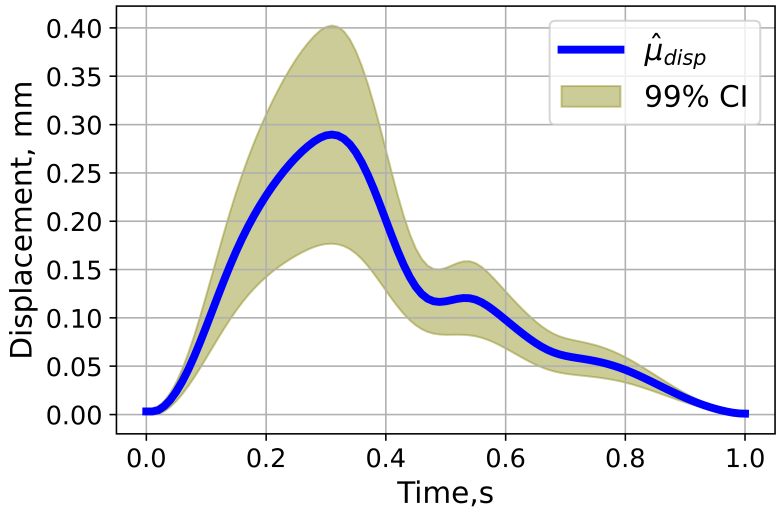

Figure 5.9: Mean response of the system along with 99% Confidence Interval calculated using the Gaussian Surrogate.

5.10 Hierarchical Bayesian Parameter Estimation

As described in detail in Section 4.5, when performing experiments, some measurements may become more similar, i.e. correlated, to each other than to other measurements. This often occurs when considering data that has a hierarchical structure like multiple measurements (e.g. performed 5 times) performed on different patients (e.g. 10 patients). Here multiple measurements will be more similar for a given patient, i.e. one patient across many days, than across

patients, i.e. one measurement compared across other patients. Similar hierarchical structures can be present in various measurement scenarios like multiple measurements of an arterial phantom (see Section 4.5 and [141]). Omitting this structure could lead to unrepresentative model predictions or inflated noise estimates. To be able to include the hierarchical information, one needs to attach an additional level to the model, a level dealing directly with the next rung in the hierarchy. It is readily understandable by observing the joint probability distribution [77]:

$$p(\Theta, \theta) = p(\theta|\Theta)p(\Theta) \tag{5.16}$$

where, in the context of this hierarchical approach, θ will denote some lower level parameter one tries to estimate (e.g. given patient's arterial stiffness) and Θ some upper level parameter (e.g. the arterial stiffness of the patient group or patients from a given city, environment etc.). Now the prior distribution has its own prior distribution, often called the hyperprior distribution [77, 92]. Moreover, a given subject-specific parameter vector (e.g. j amount of Young's moduli), θ_j, is dependent on the group level parameter (e.g. a group Young's modulus), Θ_i, such that each group level parameter is jointly distributed with each subject parameter, $p(\theta_j, \Theta)$. The notation shows that the relationship is induced through conditional probability. It is also obvious that the model is extensible to include more hierarchies. Further assuming the property of exchangeability [77], a general hierarchical specification of the posterior may be written as follows:

$$p(\theta_j, \Theta|y) \propto p(y_{i,j}|\theta_j, \Theta) \cdot p(\theta_j, \Theta) = p(y_{i,j}|\theta_j) \cdot p(\theta_j|\Theta) \cdot p(\Theta) \tag{5.17}$$

where $\theta \in \mathrm{R}^{nx1}$, $\Theta \in \mathrm{R}^{mx1}$ are both vectors of parameters of the subject level prior and group level prior (hyperprior) respectively. The whole model can be best visualized graphically, using a Kruschke-style diagram [92], see Figure 5.10. The dependencies present in the graph can also be expressed mathematically as follows:

$$
\begin{aligned}
y_{i|j} &\sim N(\mu_{i|j}, \sigma_n) \\
\mu_{i|j} &= GP(t_i, E_j) \\
\sigma_n &\sim U(L_n, H_n) \\
E_j &\sim N(\mu_E, \sigma_E) \\
\mu_E &\sim N(M_E, S_E) \\
\sigma_E &\sim U(L_E, H_E)
\end{aligned}
\tag{5.18}
$$

For both the graph and the equations the interpretation is the same. Firstly, the data are modeled as coming from a normal distribution, with a mean for each time point, t_i for a given subject j. It is important to note that variance on the data level has no dependence on i or j, this explicitly means that the same underlying variance, so-called noise variance, is applied for all the data points. The mean, $\mu_{i|j}$ is calculated similarly to how one would do for Bayesian Linear Regression: except for the (generalized) linear model, a Gaussian Process is responsible for providing the predictions. This Gaussian Process further depends on E_j, introduced in Section 5.1. Subsequently, an upper-level, group prior (hyperprior) distribution is placed on this parameter. The hyperprior is given fixed parameter values as inputs. The choice to set up the problem to estimate Young's modulus (and its related parameters) has already been justified previously, see Section 5.9.

5.10.1 Parameter estimation using SGPR within MCMC

The whole Bayesian framework used within this work was developed within a Bayesian modeling Python library PYMC, version 5.10.3 [176], while some of the presented visualizations were facilitated using the Python library for (exploratory) Bayesian Analyses, Arviz version 0.17.0 [177]. The model structure was the same as that provided in Equation 5.18 and image 5.10. The Gaussian Process was defined in a separate custom operation and then called from within the PYMC model definition. To make the high-dimensional parameter space more amenable to exploration, the priors and hyperpriors were standardized. [176] A regular Metropolis sampler was assigned for the individual Young's moduli, while the rest was explored using the No U-Turn Sampler (NUTS). The analysis was set up to run on 4 chains with 3000 samples of burn-in and 10 000 samples of regular exploration (in total yielding 52 000 samples). The burn-in period is necessary as it is the moment when the chain begins to explore

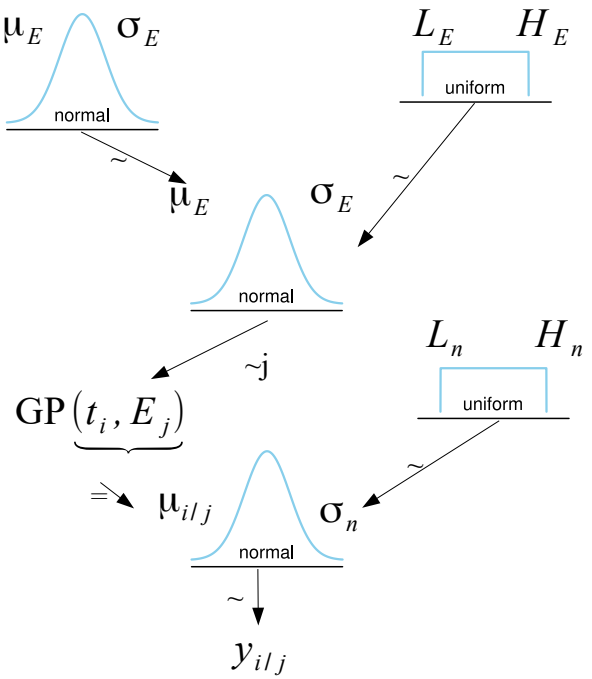

Figure 5.10: A Kruschke-style diagram [92] depicting the overall structure of the hierarchical model. From the bottom up: datapoints, y_i are nested within a subject, j. The data are subject to normal likelihood with the subject dependent mean $\mu_{i|j}$ and standard deviation σ_n. The noise is modeled as a uniform random variable. The mean is modeled by a time-dependent Gaussian Process, where E_j is parametrized to be subject specific the Young's modulus. It in turn has a prior distribution that is modeled by a normal distribution with mean μ_E and standard deviation σ_E. They have their respective (normal and uniform) hyperpriors.

the space and auto-tunes other hyperparameters related directly to sampling (e.g. NUTS' leapfrog step). [92] Large sampling sizes are the norm, not the exception, for Bayesian modeling and would require large resources to evaluate the full 3D FSI model around 52 000 times to complete the presented analysis. Figure 5.11 presents the most important results. Figure 5.11 **A** depicts the estimated values of parameters, which are the MAP (MAP stands for *maximum*

a posteriori estimate and is the mode of the posterior distribution, presented with magenta circles) and the mean (green squares) along with 99% HDI of the distribution. Highest Density Interval, HDI, is the region where most of the probability density is concentrated, in this case, 99% of it. The narrower it is the more certain the estimate. Given that, as opposed to frequentist statistical analyses, the Bayesian analysis yields full (sampled) distributions, the distributional summaries tend to be more intuitive and robust. [92]. The estimates (mean and MAP) are plotted next to the true values (red crosses) and randomly initialized values (blue triangles). All estimates were obtained from the pooled samples of the four chains. As can be seen most of the estimated values are either very close to the true values or the true values are covered by the uncertainties superimposed on the estimates. For individual level parameters (E[1] - E[10]) both of the estimates are very close to the true values and the uncertainties represented by the 99% HDI are tight. For the group level parameters, μ_E and σ_E the uncertainties are a bit wider, which is due to relatively large uncertainty placed on the group mean and the low count of the individuals representative of it and not due to the quality of the chain itself (see the paragraph below). The exact values of all estimates, along with their initialized values (again, which were randomized) and the true values based on distributions derived from experimental and literature data are collected in Table 5.4.

Figures 5.11 **B** and **C** show the 100 posterior predictions for the measured displacements using the first and fifth estimates of the Young's moduli (the choice which to show was random). Such a posterior predictive check is carried out by placing the posterior values of parameters as the new definitions of respective distribution. In the next step, the data are sampled, and the model is run (in this case Gaussian Process). The predicted displacements closely match the true ones and the noise is filtered out. Figure 5.11 **C** shows that the calculated error metrics for the predictive performance of the model (same as in section 5.3) are very low. The plotted error metrics are averaged first across a given displacement curve (i.e. time wise) and then across 100 posterior predictions. Root Mean Squared Error (RMSE) and Maximum Error (MaxErr) are considered to be 'worst case scenario' metrics, as they tend to ascribe large weighting to outliers or uncommon values (particularly maximum error), however here can be seen that the two worst predictive performances are for $\Delta d(E[2])$ at about ~ 0.004266 mm for RMSE and for $\Delta d(E[9])$ at around ~ 0.01255 mm for MaxErr. Most of the time the metrics do not exceed 10% of the average displacement and generally are very well below 5%. The numerical summary of

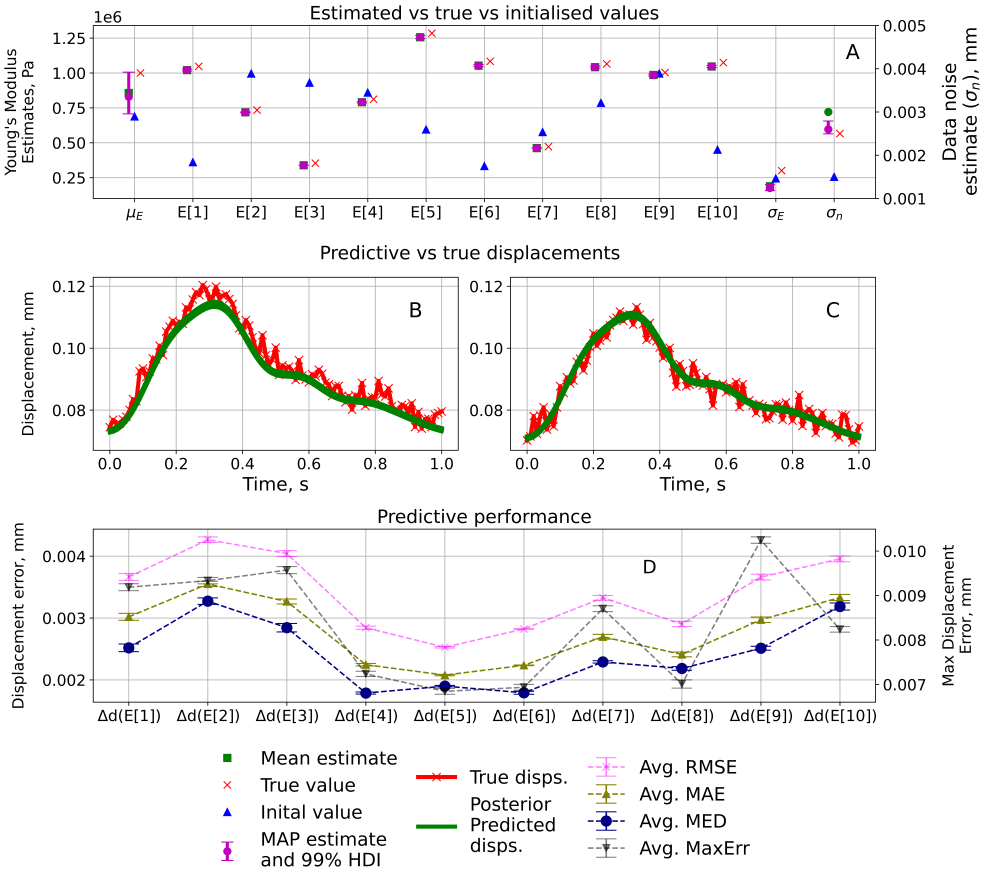

Figure 5.11: Summary of the results of the Bayesian Analysis. Figure **A** shows the inversely estimated parameters (MAP and mean, both averaged across chains) with a 99% HDI superimposed against the true values and values at which the chains were (randomly) initialized. **Note:** For readability the parameter estimates of an individual Young's modulus are described generically as $E[x]$ rather than $\hat{\mu}_{E[x]}$ or $\hat{E}[x]_{MAP}$ for mean and MAP respectively. Figures **B** and **C** show the true data using which the Moduli were estimated along with 100 posterior predictions from estimated distributions. Figure **D** shows the median Gaussian Process prediction errors calculated using RMSE, MAE, and MED for each individual Young's Modulus with a 99% HDI superimposed. On average the relative errors are around 2.5%, going up to 10% when considering the extreme edges of the HDIs (i.e. $0.0075mm$ and beyond).

posterior model predictions along with error metrics and relative error metrics is presented in Table 5.10.1. As can be seen, the errors do not tend to exceed the 5% threshold (even for RMSE which tends to be very sensitive), but for the Max Error metric which only once exceeds the 10% threshold (i.e. 11.183% for $\Delta d(E[9])$). The very narrow 99% Credible Interval estimates for errors also demonstrate very low uncertainty present in the error estimates.

Table 5.3: Summary table of all the time- and sample-averaged posterior predictive errors. Most samples do not exceed relative errors 5%, save for worst-case-scenario Max Errors which exceeds 10% for one of the cases but generally stay below this limit. **Note:** All quantities have the units of millimeters except for relative errors quoted in percent.

Displacement	**Mean RMSE**	**Mean rel. RMSE**	**CI** 0.5% **RMSE**	**CI** 99.5% **RMSE**	**Mean MAE MAE**	**Mean rel. MAE**	**CI** 0.5% **MAE**	**CI** 99.5% **MAE**
$\Delta d(E[1])$	3.664E-03	3.92	3.608E-03	3.720E-03	3.019E-03	3.229	2.964E-03	3.075E-03
$\Delta d(E[2])$	4.266E-03	3.23	4.221E-03	4.310E-03	3.551E-03	2.686	3.507E-03	3.594E-03
$\Delta d(E[3])$	4.042E-03	1.49	3.995E-03	4.088E-03	3.268E-03	1.207	3.228E-03	3.307E-03
$\Delta d(E[4])$	2.843E-03	2.41	2.817E-03	2.869E-03	2.246E-03	1.901	2.226E-03	2.265E-03
$\Delta d(E[5])$	2.527E-03	3.41	2.516E-03	2.538E-03	2.077E-03	2.806	2.067E-03	2.088E-03
$\Delta d(E[6])$	2.823E-03	3.21	2.818E-03	2.828E-03	2.236E-03	2.540	2.230E-03	2.241E-03
$\Delta d(E[7])$	3.326E-03	1.67	3.288E-03	3.365E-03	2.697E-03	1.354	2.658E-03	2.736E-03
$\Delta d(E[8])$	2.904E-03	3.32	2.862E-03	2.945E-03	2.411E-03	2.760	2.373E-03	2.450E-03
$\Delta d(E[9])$	3.662E-03	4.00	3.615E-03	3.708E-03	2.972E-03	3.243	2.928E-03	3.015E-03
$\Delta d(E[10])$	3.956E-03	4.63	3.908E-03	4.004E-03	3.332E-03	3.896	3.282E-03	3.381E-03

Displacement	**Mean MEDS**	**Mean rel. MEDS**	**CI** 0.5% **MEDS**	**CI** 99.5% **MEDS**	**Mean MAXERR**	**Mean rel. MAXERR**	**CI** 0.5% **MAXERR**	**CI** 99.5% **MAXERR**
$\Delta d(E[1])$	2.518E-03	2.693	2.456E-03	2.580E-03	9.186E-03	9.823	9.108E-03	9.265E-03
$\Delta d(E[2])$	3.273E-03	2.476	3.220E-03	3.325E-03	9.334E-03	7.061	9.263E-03	9.406E-03
$\Delta d(E[3])$	2.844E-03	1.050	2.776E-03	2.912E-03	9.571E-03	3.535	9.491E-03	9.650E-03
$\Delta d(E[4])$	1.787E-03	1.513	1.771E-03	1.804E-03	7.237E-03	6.125	7.178E-03	7.296E-03
$\Delta d(E[5])$	1.899E-03	2.565	1.888E-03	1.911E-03	6.844E-03	9.244	6.779E-03	6.910E-03
$\Delta d(E[6])$	1.792E-03	2.036	1.768E-03	1.815E-03	6.935E-03	7.879	6.870E-03	7.000E-03
$\Delta d(E[7])$	2.290E-03	1.150	2.272E-03	2.309E-03	8.703E-03	4.368	8.637E-03	8.769E-03
$\Delta d(E[8])$	2.185E-03	2.501	2.155E-03	2.214E-03	7.010E-03	8.023	6.917E-03	7.103E-03
$\Delta d(E[9])$	2.512E-03	2.741	2.480E-03	2.544E-03	1.025E-02	11.183	1.017E-02	1.032E-02
$\Delta d(E[10])$	3.186E-03	3.726	3.129E-03	3.243E-03	8.239E-03	9.635	8.176E-03	8.302E-03

Besides looking at the comparison of estimated and true values and assessing the predictive performance of the model and its ability, the diagnostic metrics that describe the inference itself should also be analyzed. Starting with $\hat{R}$ (see Section 3.6.2). It is recommended to keep its value below 1.05, otherwise, the chains should be run for a longer time [77]. In Table 5.4 all of $\hat{R}$ statistics lie below 1.05, even below or on the level of the stricter condition of $\hat{R} \leq 1.01$. This indicates that any errors in estimation would most likely not be solved by running the simulations any longer (i.e. drawing more samples would not improve the quality of estimates). Bayesian analyses, especially employing MCMC (and some with HMC), tend to produce somewhat autocorrelated chains, rendering the produced samples not quite independent and thus unsuited for inference. Thus one of the most important diagnostics is the Effective Sample Size (ESS, see section 3.6.2). Having a large ESS enables one to have a more robust and trustworthy estimates of the parameters. The recommended value for ESS is above 100, ideally above 200 [178, 179], however generally the larger ESS the more trustworthy the estimates. In Table 5.4 there are two quantities related to ESS, i.e. ESS_{bulk} and ESS_{tail}. They originate from a similar principle of ESS introduced in Section 3.6.2, but are responsible for 'enough samples' of two different estimates. Large ESS_{bulk} serve as a basis for accurate group-tendency estimates (i.e. mean, median, mode etc.), while large ESS_{tail} similarly are responsible for distribution-tail related summaries (quartiles, quantiles, intervals, HDIs etc.). [178] Both of these statistics lie on the safe side, being well over 1000 for most of the parameters, over 10 000 for the group-level mean estimate (for ESS_{bulk}), and well above 100 for two of the parameters (E[7] and E[8]). Finally for MCSE applied to means and variances, one can see that due to large ESS they have been shrunk to small values, in some cases near zero (reported as 0 in Table). These low values of this statistic indicate a relatively large certainty in the estimates produced by the chains, i.e. the standard error on the mean and variance (standard deviation in this case) are low. [140] The quality of the developed model along with the low error values produced by the model with reference to test data, makes it evident that the whole procedure does a very good job of estimating the arterial stiffness.

Table 5.4: Summary table of statistics related to both Bayesian estimates as well as the estimation procedure itself.

Param.	$\hat{\mu}$	$\hat{\sigma}$	**MAPs**	θ_{true}	**Init.**	**HDI 0.5%**
μ_E	8.56E+05	5.82E+04	8.29E+05	1.00E+06	6.84E+05	7.06E+05
E[0]	1.02E+06	3.00E+03	1.02E+06	1.04E+06	9.95E+05	1.01E+06
E[1]	7.18E+05	1.40E+03	7.18E+05	7.34E+05	9.30E+05	7.15E+05
E[2]	3.39E+05	4.00E+02	3.39E+05	3.54E+05	8.593E+05	3.38E+05
E[3]	7.90E+05	1.80E+03	7.91E+05	8.12E+05	5.95E+05	7.87E+05
E[4]	1.26E+06	4.40E+03	1.26E+06	1.28E+06	3.34E+05	1.25E+06
E[5]	1.05E+06	3.00E+03	1.05E+06	1.08E+06	5.77E+05	1.05E+06
E[6]	4.67E+05	6.00E+02	4.62E+05	4.72E+05	7.86E+05	4.61E+05
E[7]	1.04E+06	3.00E+03	1.04E+06	1.07E+06	9.96E+05	1.04E+06
E[8]	9.85E+05	2.60E+03	9.85E+05	1.00E+06	4.50E+05	9.81E+05
E[9]	1.05E+06	3.00E+03	1.05E+06	1.07E+06	3.61E+05	1.04E+06
σ_E	1.88E+05	1.00E+04	1.76E+05	3.00E+05	2.44E+05	1.57E+05
σ_n	3.00E-03	0.00E+00	2.60E-03	2.50E-03	1.50E-03	2.50E-03

Param.	**HDI 99.5%**	$MCSE[\hat{\mu}]$	$MCSE[\hat{\sigma}]$	ESS_{bulk}	ESS_{tail}	$\hat{R}$
μ_E	1.00E+06	6.00E+02	4.00E+02	1.05E+04	8.77E+03	1.00
E[0]	1.02E+06	0.00E+00	0.00E+00	1.27E+03	1.00E+03	1.01
E[1]	7.20E+05	0.00E+00	0.00E+00	1.07E+03	1.13E+03	1.01
E[2]	3.40E+05	0.00E+00	0.00E+00	1.45E+03	1.66E+03	1.00
E[3]	7.94E+05	0.00E+00	0.00E+00	1.45E+03	1.78E+03	1.00
E[4]	1.26E+06	2.00E+02	0.00E+00	1.34E+03	1.08E+03	1.00
E[5]	1.06E+06	0.00E+00	0.00E+00	1.65E+03	1.49E+03	1.01
E[6]	4.63E+05	0.00E+00	0.00E+00	1.27E+03	1.23E+03	1.00
E[7]	4.61E+05	0.00E+00	0.00E+00	1.01E+03	7.60E+02	1.00
E[8]	9.91E+05	0.00E+00	0.00E+00	9.81E+02	1.01E+03	1.00
E[9]	1.05E+06	0.00E+00	0.00E+00	1.05E+03	1.32E+03	1.01
σ_E	2.00E+05	2.00E+02	0.00E+00	5.93E+03	4.22E+03	1.00
σ_n	2.79E-03	0.00E+00	0.00E+00	9.43E+03	8.67E+03	1.00

6

Conclusions and further research

The primary motivation of this work was to develop a method that would be an alternative to modern, mostly absent, approaches to local arterial stiffness estimation. [5] The upper hand of this method over the others would come into play in a few important aspects. First, the method would measure local arterial stiffness, which is one of the best early indicators of the onset of arterial pathologies [6]. Second, it would implement the ultrasound system which is easy to operate by medical personnel and highly cost-effective. Third, the method would provide uncertainty bands for the estimates, which would help the physician's decision making process. [1] The crucial part of the developed method was the surrogate model for fast and accurate emulation of a verified and validated Fluid-Structure Interaction model of a carotid artery phantom. Without this surrogate, the most important task, i.e. parameter estimation, would be impossible as computation times for Full Order Models, like FSI, are too prohibitive. As stated in the *Objectives and Methodology* Section of the Introduction, the development of this method is a fairly involved, multi-step procedure. The results and insights gained from each of the taken steps, are discussed briefly below, along with suggested possible avenues of future research.

The application sections of this work started by first constructing an experimental rig for carrying out pressure, flow, camera and ultrasound measurements. The object of interest was a phantom model of the left common carotid artery. The measurements were carried out to be able to generate Boundary and Initial Conditions for the FSI model as well as to be able to validate the FSI model itself. It was achieved by developing an in-house control application in LabVIEW and subsequently constructing the post-processing algorithms employed to analyze the pressure, flow and image raw data resulting from a series of experiments over physiological ranges of pressures.

Another important step was a statistical analysis of measurement variability, carried out in order to be able to accurately characterize the uncertainties about the data, which was key to explain away the noise present due to various factors and leave only the variability present due to equipment construction and operation. The method applied for modeling the data and its uncertainty, as mentioned in section 4.5 was Mixed-Effects modeling. The hierarchy of this model used, from bottom up, was: variation due to equipment noise, variation due to cycle to cycle variation (nested in experiments) and experiment to experiment variation (belonging to a given equipment, i.e. flowmeter). The results of this analysis proved the experiments to be highly reproducible. The residual variation σ_{resid} seemed consistent with high frequency noise from uncontrollable sources (environmental and equipment noise). In comparison with the data originating from the camera, the ultrasound derived displacements exhibit good agreement (mean difference of 0.0113 mm), although slightly more deviation on average between ultrasound and camera than between the two cameras (see Table 4.6). The agreement of ultrasound derived displacements with camera-measured displacements, supports further use of ultrasound as a method for describing the displacement of arterial walls. The experimental phantom and ultrasound measurement system have thus been confirmed to provide a somewhat consistent experimental model for generating data for developing and testing new methods for non-invasive assessment of the common carotid artery. There are a few possible extensions of this methodology that could be researched further. First, the implemented hierarchy could be further extended to treat the equipment itself (transducers, flowmeters, cameras) as themselves being nested in a 'general' transducer, which could more clearly shed light on present equipment-to-equipment (e.g. transducer 1 to transducer 2) variation as well as on the variance of the product performance itself. Such variation could arise due to different equipment being tuned slightly different, their different spatial location, manufacturing variance and other factors. Such

model was entertained, but due to time constraints and instability issues remained only at a proposal stage. Furthermore the models could be rephrased in a Bayesian context, which, as mentioned, in Section 3.5, is a far more interpretable and robust framework for data analysis. However, the sheer volume of data resulting from amount of equipment (2 cameras, 8 transducers and 2 flowmeters), measurements (4 series, 7 measurements each, hundreds of cycles per measurement), when placed in a Hierarchical Bayesian context to reflect the LME structure, yielded unreasonably long computational times and highly autocorrelated chains. [92] Since such an approach could provide a lot more insight into the data, it is heavily recommended to be explored in further research. Naturally, application of other approaches and algorithms would be welcome to carry out the error analysis. Gaussian Error Propagation is a very commonly applied approach for analysing uncertainty present within physical experiments [180]. Its primary drawback however is the fact that it assumes a known functional form with which to calculate derivatives for variables of interest. In the case of this work, coming up with such derivatives and their form was unclear, and thus it was left out. However it would certainly be an interesting way to quantify the uncertainty of the measurements, as the end result would allow to properly describe the relationships present between the system output (phantom displacements) and input variables (pressure, flow, etc. and their measuring devices' related uncertainties) analytically.

When it comes to the Fluid-Structure interaction models, the modeled bodies were reflected well using the basic hyperelastic and, to some degree, linear elastic models. One thing that could directly (but rather not dramatically) improve the faithfulness model of the phantom would be to perform radial tensile material tests on it. By placing it in the measuring site (the aquarium), fixing it and slowly ramping up the pressure while measuring the displacements, one could obtain a strain-stress curve. This would be a valuable resource, as the literature is lacking in such a characteristic for most materials. This could be taken further, as by having a much clearer characteristic of the material, one could be tempted to use a bit more exact models like Mooney-Rivlin or even, if real arteries were measured, Holzapfel-Ogden model. This, of course, raises another possible extension of the model - extension to real human patient data and not arterial phantoms. Although the groundwork for such an application has been laid within this work, it was ultimately the lack of time and the ill timing of the data from the first patients that led to them being not analyzed and ultimately excluded from this work.

This work also addresses the surrogate model development in-depth and describes the modern challenges of constructing such a model. From experimental design point of view, rather than employing conventional techniques, like simple random sampling and Latin-Hypercube sampling, this study utilized quasi-random sequences. Time and time again, they prove to be the superior choice due to their better space coverage and faster asymptotic convergence of estimated statistical moments [71, 74]. One interesting possible extension of experimental design is Sequential Sampling. [181] There instead of generating N amount of samples and then running the models, sequential sampling employs the minimization approach. Simply put, it works by looking for regions in the experimental space, which when sampled, could reduce the uncertainty about the model output the most. While this approach was implemented, due to time constraints it has been dropped.

For the surrogate model itself, the two considered models were the (Sparse) Gaussian Process surrogate and the SVD-based Reduced Order Model. Starting with the latter, many other ideas were entertained such as a (kernelized) Dynamic Mode Decomposition Model [182], Gaussian Process Latent Variable Models [183] and others. They did not provide superior performance to the ROM (developed in Section 5.4) and thus they have been dropped from this work. However, it is almost certain that by further tuning them (and with more data) they would be able to outperform the considered ROM. The former of the two applied models, the Gaussian Process Regression surrogate (developed in Section 5.2), was the choice for the following tasks, due to its superior predictive performance over the Reduced Order Model. This dissertation clearly demonstrated that by replacing the FSI model with a non-parametric SGPR surrogate, trained on a combination of experimental and literature data, it is possible to approximate the FSI output fairly well and then apply the surrogate with much shorter evaluation tasks. It has been mentioned in Section 3.2.1, that Gaussian Process Regression models are highly sensitive towards kernel choice. Thus, to gain optimal performance, they were tuned manually taking into account the underlying overall shape of the model response and the believed interaction of input variables. However, when applying the methodology in commercial setting, or in general for future research, it would be interesting to see a development of a method for automatic kernel selection along with its combination to best fit the data. Although such approaches have been attempted recently [184], it is still a very active area of research in the Gaussian Process community.

Following from the Section *Objectives and Methodology*, next item on the to-do list was the Uncertainty Quantification and Sensitivity Analysis of the Full Order FSI model. This has been done for two reasons: first to indicate the future areas of research when it comes to narrowing down the uncertainty on the considered variables and parameters. Second, it allowed to fix some of the considered variables with the lowest contribution to the overall model variance. Within this work, the Monte Carlo-based Global Uncertainty Quantification and Sensitivity Analysis (UQSA) have been applied. In this approach sensitivity is quantified using the Sobol Indices, which are approximated using Monte Carlo estimators. These estimators need to converge, i.e. stabilize, which often requires numerous evaluations [71]. Without employing the GP emulator, both stabilizing the indices and acquiring their uncertainty estimates (achieved through bootstrapping, see [133]) would require considerable time. The conducted analyses revealed that, given the existing assumptions, Young's modulus most significantly impacts the output variance of the FSI Neo-Hookean and linear elastic-based models, making it the most critical factor. Consequently, it is essential to further explore Young's modulus to refine its prior distribution (marked by a high First Order index), while it may be appropriate to maintain all other variables at their standard levels for different model applications (marked by a high Total Sobol index). Naturally, were the underlying material model extended, its parameters may require a separate Sensitivity Analysis. [174] It is important to note that some variables exhibited correlations and therefore needed to be decorrelated during the analysis, leading to the creation of both Full and Marginal Sobol indices. However, these factors still represented a minor portion of the overall variance. This method could be adapted to encompass correlations among more than two variables and, by understanding their interdependencies, potentially incorporate discrete (e.g. age) or binary (e.g. gender) variables, thereby generating significantly more intricate relationships (joint distribution) compared to (multivariate) normal distributions.

Finally, the surrogate was employed in another highly computationally intensive context, specifically for parameter estimation via Hierarchical Bayesian Regression. The objective was to determine, from displacement data, the individual levels of Young's modulus and the overarching distribution at the group level that controls the individual members. This approach concurrently estimates characteristics specific to each subject while also inferring the parameters that regulate entire groups or populations. Bayesian modeling, known for providing extensive data for inference, is notably demanding in terms of computation. In this study, approximately 50,000 samples were required to achieve reliable

parameter distributions for more than ten subjects, distributed among four chains. By embedding the Gaussian Process surrogate within the Bayesian inference machine, it was possible to estimate the desired parameters with high accuracy. As a result, the uncertainty associated with group-level parameters could be estimated, which typically requires much more than 10 subjects. The analysis is readily extensible to many more subjects (although at additional computational cost), and more importantly, it is possible to extend the model to many more parameters with possible correlations. Despite the efforts with this concept, it resulted in unstable and highly correlated chains (even with standardization and reparametrization [185]) and inaccurate estimates, likely due to the negligible impact of the remaining parameters, pushing the entire estimation process towards non-identifiability. Despite these issues, this approach remains a promising field for future investigation.

To summarize, this dissertation revolved primarily around the development of a novel method for non-invasive estimation of local arterial stiffness and validating its constitutents on arterial phantom data. The method development was built out of a few components: development of the experimental rig served to validate the ultrasound as a reliable instrument for accurate displacement detection. The developed linear mixed-effects models served as a way to determine whether the experiments were reproducible by estimating the uncertainty present within them. The experiments also doubled as a tool for dataset generation for the purposes of development of the Fluid-Structure Interaction model and surrogate development. The FSI model was developed and validated, but in itself proved to be prohibitive when it comes to implementing it in computationally demanding tasks, such as parameter estimation. Thus, the surrogate was developed and would form the backbone of the developed method, whose goal was to serve as an alternative for current arterial stiffness estimation procedures. The use of a Gaussian Process Regression surrogate was motivated by an extensive literature review (see Introduction), but by no means is it the only choice as among other popularly applied modes were Support Vector Machines, Neural Networks, classical Multiple Linear Regression and many others. The developed surrogate was applied to tasks of Uncertainty Quantification and Sensitivity Analysis to allow to comment on the quality of the FSI model and to suggest further research developments. The methodology loop is closed with the application of the Gaussian surrogate to parameter estimation of the arterial stiffness, which was done employing Hierarchical Bayesian estimation.

Bibliography

[1] Mariachiara Di Cesare, Honor Bixby, Thomas Gaziano, Lisa Hadeed, Chodziwadziwa Kabudula, Diana Vaca McGhie, Jeremiah Mwangi, Borjana Pervan, Pablo Perel, Daniel Piñeiro, Sean Taylor, and Fausto Pinto. World heart report 2023: Confronting the world's number one killer. *WHF*, 2023.

[2] Shanthi Mendis, Pekka Puska, B editors Norrving, World Health Organization, et al. *Global atlas on cardiovascular disease prevention and control*. World Health Organization, 2011.

[3] Najeeb A Shirwany and Ming-hui Zou. Arterial stiffness: a brief review. *Acta Pharmacologica Sinica*, 31(10):1267–1276, 2010.

[4] Seamus P Whelton, Ron Blankstein, Mouaz H Al-Mallah, Joao AC Lima, David A Bluemke, W Gregory Hundley, Joseph F Polak, Roger S Blumenthal, Khurram Nasir, and Michael J Blaha. Association of resting heart rate with carotid and aortic arterial stiffness: multi-ethnic study of atherosclerosis. *Hypertension*, 62(3):477–484, 2013.

[5] Patrick Segers, Ernst R Rietzschel, and Julio A Chirinos. How to measure arterial stiffness in humans. *Arteriosclerosis, thrombosis, and vascular biology*, 40(5):1034–1043, 2020.

[6] Tânia Pereira, Carlos Correia, and Joao Cardoso. Novel methods for pulse wave velocity measurement. *Journal of medical and biological engineering*, 35(5):555–565, 2015.

[7] Non-invasive in-vivo assessment of the stiffness of human artery walls. `https://enthral.pl/`, August 2022.

[8] Richard C Aster, Brian Borchers, and Clifford H Thurber. *Parameter estimation and inverse problems*. Elsevier, 2018.

[9] CJ Reddy. *Finite element method for eigenvalue problems in electromagnetics*, volume 3485. NASA, Langley Research Center, 1994.

[10] J.D. Anderson. *Computational Fluid Dynamics: The Basics with Applications*. McGraw-Hill International Editions: Mechanical Engineering. McGraw-Hill, 1995.

[11] Matthew Hirschhorn, Vakhtang Tchantchaleishvili, Randy Stevens, Joseph Rossano, and Amy Throckmorton. Fluid–structure interaction modeling in cardiovascular medicine–a systematic review 2017–2019. *Medical engineering & physics*, 78:1–13, 2020.

[12] Steve A Maas, Gerard A Ateshian, and Jeffrey A Weiss. Febio: History and advances. *Annual review of biomedical engineering*, 19:279–299, 2017.

[13] Ansys Inc. Ansys Documentation, 2021.

[14] DY Dhande and DW Pande. A two-way fsi analysis of multiphase flow in hydrodynamic journal bearing with cavitation. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 39:3399–3412, 2017.

[15] Mobin Masoomi and Amir Mosavi. The one-way fsi method based on rans-fem for the open water test of a marine propeller at the different loading conditions. *Journal of Marine Science and Engineering*, 9(4):351, 2021.

[16] Mikel Ezkurra, Jon Ander Esnaola, Manex Martinez Agirre, Unai Lertxundi, and Unai Etxeberria. Analysis of one-way and two-way fsi approaches to characterise the flow regime and the mechanical behaviour during closing manoeuvring operation of a butterfly valve. *World Academy of Science, Engineering and Technology, International Science Index 136, International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering*, 12:409–415, 2018.

[17] Gene Hou, Jin Wang, and Anita Layton. Numerical methods for fluid-structure interaction—a review. *Communications in Computational Physics*, 12(2):337–377, 2012.

[18] Y Pratumwal, W Limtrakarn, and P Premvaranon. The analysis of blood flow past carotid bifurcation by using the one-way fluid solid interaction technique (fsi). *Journal of Research and Applications in Mechanical Engineering*, 2(1):57–64, 2014.

[19] Alex G Kuchumov, Vasily Vedeneev, Vladimir Samartsev, Aleksandr Khairulin, and Oleg Ivanov. Patient-specific fluid–structure interaction model of bile flow: comparison between 1-way and 2-way algorithms. *Computer Methods in Biomechanics and Biomedical Engineering*, 24(15):1693–1717, 2021.

[20] Andrzej Polanczyk, Michal Podgorski, Maciej Polanczyk, Natasha Veshkina, Ireneusz Zbicinski, Ludomir Stefanczyk, and Christoph Neumayer. A novel method for describing biomechanical properties of the aortic wall based on the three-dimensional fluid-structure interaction model. *Interactive cardiovascular and thoracic surgery*, 28(2):306–315, 2019.

[21] Vincent Mendez, Marzio Di Giuseppe, and Salvatore Pasta. Comparison of hemodynamic and structural indices of ascending thoracic aortic aneurysm as predicted by 2-way fsi, cfd rigid wall simulation and patient-specific displacement-based fea. *Computers in biology and medicine*, 100:221–229, 2018.

[22] Mo Yin Zhao, Tracie J Barber, Peter A Cistulli, Kate Sutherland, and Gary Rosengarten. Using two-way fluid-structure interaction to study the collapse of the upper airway of osa patients. *Applied Mechanics and Materials*, 553:275–280, 2014.

[23] Xiaochen Wang, Mergen H Ghayesh, Andrei Kotousov, Anthony C Zander, Marco Amabili, Joseph A Dawson, and Peter J Psaltis. Biomechanics of abdominal aortic aneurysm in the framework of windkessel effect and fully-developed inflow velocity via two-way non-linear fsi. *International Journal of Non-Linear Mechanics*, 157:104517, 2023.

[24] Ahmed Tasnub Takaddus and Abhilash J Chandy. A three-dimensional (3d) two-way coupled fluid-structure interaction (fsi) study of peristaltic flow in obstructed ureters. *International Journal for Numerical Methods in Biomedical Engineering*, 34(10):e3122, 2018.

[25] Ghalib Taimuri, Tommi Mikkola, Pentti Kujala, and Spyros Hirdaris. A two-way coupled fsi ship hard grounding dynamics model. In *IOP Conference Series: Materials Science and Engineering*, volume 1288, page 012014. IOP Publishing, 2023.

[26] Hamid Matin Nikoo, Kaiming Bi, and Hong Hao. Effectiveness of using pipe-in-pipe (pip) concept to reduce vortex-induced vibrations (viv):

Three-dimensional two-way fsi analysis. *Ocean Engineering*, 148:263–276, 2018.

[27] Eric Pedrol, Jaume Massons, Francesc Díaz, and Magdalena Aguiló. Two-way coupling fluid-structure interaction (fsi) approach to inertial focusing dynamics under dean flow patterns in asymmetric serpentines. *Fluids*, 3(3):62, 2018.

[28] S Hashemi, T Armaghani, and R Ghasemiasl. 3d numerical study of al2o3-water nanofluid unsteady natural convection via using oscillating plate by two-way fsi analysis. *Journal of Thermal Analysis and Calorimetry*, pages 1–16, 2024.

[29] Tahereh Liaghat, Francois Guibault, L Allenbach, and B Nennemann. Two-way fluid-structure coupling in vibration and damping analysis of an oscillating hydrofoil. In *ASME International Mechanical Engineering Congress and Exposition*, volume 46476, page V04AT04A073. American Society of Mechanical Engineers, 2014.

[30] AK Khe, AA Cherevko, AP Chupakhin, MS Bobkova, AL Krivoshapkin, and K Yu Orlov. Haemodynamics of giant cerebral aneurysm: A comparison between the rigid-wall, one-way and two-way fsi models. In *Journal of Physics: Conference Series*, volume 722, page 012042. IOP Publishing, 2016.

[31] Bijan Mohammadi and Olivier Pironneau. Shape optimization in fluid mechanics. *Annu. Rev. Fluid Mech.*, 36:255–279, 2004.

[32] Kevin N Otto and Erik K Antonsson. Tuning parameters in engineering design. *Journal of Mechanical Design*, 115:14–19, 1993.

[33] S Wojtkiewicz, M Eldred, R Field, Jr, A Urbina, and J Red-Horse. Uncertainty quantification in large computational engineering models. In *19th AIAA Applied Aerodynamics Conference*, page 1455, 2001.

[34] Huibin Liu, Wei Chen, and Agus Sudjianto. Relative entropy based method for probabilistic sensitivity analysis in engineering design. *Journal of Mechanical Design*, 128(2):326–336, 2006.

[35] Felipe AC Viana, Christian Gogu, and Tushar Goel. Surrogate modeling: tricks that endured the test of time and some recent developpilbeam c.ments. *Structural and Multidisciplinary Optimization*, pages 1–28, 2021.

[36] D. Denyer and C. Pilbeam. Doing a literature review in business and management [presentation to the british academy of management doctoral symposium]. `https://www.ifm.eng.cam.ac.uk/uploads/Research/RCDP/Resources/Working_with_literature_for_Cambridge.pdf`, 2010. Accessed: 2023-01-31.

[37] Fang Li, Mihkel Kõrgesaar, Pentti Kujala, and Floris Goerlandt. Finite element based meta-modeling of ship-ice interaction at shoulder and midship areas for ship performance simulation. *Marine Structures*, 71:102736, 2020.

[38] Tao Chen, Kunn Hadinoto, Wenjin Yan, and Yifei Ma. Efficient metamodelling of complex process simulations with time–space-dependent outputs. *Computers & chemical engineering*, 35(3):502–509, 2011.

[39] Christopher S Thurman and J Ryan Somero. Comparison of metamodeling methodologies through the statistical-empirical prediction modeling of hydrodynamic bodies. *Ocean Engineering*, 210:107566, 2020.

[40] José Antonio Expósito Carrillo, Francisco José Sánchez de La Flor, and José Manuel Salmerón Lissén. Single-phase ejector geometry optimisation by means of a multi-objective evolutionary algorithm and a surrogate cfd model. *Energy*, 164:46–64, 2018.

[41] Yoke Yuan Loy, Gade Pandu Rangaiah, and Samavedham Lakshminarayanan. Surrogate modelling for enhancing consequence analysis based on computational fluid dynamics. *Journal of Loss Prevention in the Process Industries*, 48:173–185, 2017.

[42] Smail Khalfallah and Adel Ghenaiet. Radial basis function-based shape optimization of centrifugal impeller using sequential sampling. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 229(4):648–665, 2015.

[43] Manuel Bompard, Jacques Peter, and Jean-Antoine Desideri. Surrogate models based on function and derivative values for aerodynamic global optimization. In *V European Conference on Computational Fluid Dynamics ECCOMAS CFD 2010*, 2010.

[44] Zebin Zhang, Bruno Demory, Manuel Henner, Pascal Ferrand, Frédéric Gillot, Youssef Beddadi, François Franquelin, and Vivien Marion. Space infill study of kriging meta-model for multi-objective optimization of an

engine cooling fan. In *Turbo Expo: Power for Land, Sea, and Air*, volume 45578, page V01AT10A003. American Society of Mechanical Engineers, 2014.

[45] Brent A Craven, Kenneth I Aycock, Luke H Herbertson, and Richard A Malinauskas. A cfd-based kriging surrogate modeling approach for predicting device-specific hemolysis power law coefficients in blood-contacting medical devices. *Biomechanics and Modeling in Mechanobiology*, 18:1005–1030, 2019.

[46] Kim Goethals, Ivo Couckuyt, Tom Dhaene, and Arnold Janssens. Sensitivity of night cooling performance to room/system design: surrogate models based on cfd. *Building and Environment*, 58:23–36, 2012.

[47] Gabriel FN Gonçalves, Assen Batchvarov, Yuyi Liu, Yuxin Liu, Lachlan R Mason, Indranil Pan, and Omar K Matar. Data-driven surrogate modeling and benchmarking for process equipment. *Data-Centric Engineering*, 1:e7, 2020.

[48] Erkan Gunpinar, Umut Can Coskun, Mustafa Ozsipahi, and Serkan Gunpinar. A generative design and drag coefficient prediction system for sedan car side silhouettes based on computational fluid dynamics. *Computer-Aided Design*, 111:65–79, 2019.

[49] Gizem Demirel, Erdem Acar, Kutay Celebioglu, and Selin Aradag. Cfd-driven surrogate-based multi-objective shape optimization of an elbow type draft tube. *International Journal of Hydrogen Energy*, 42(28):17601–17610, 2017.

[50] Romit Maulik, Himanshu Sharma, Saumil Patel, Bethany Lusch, and Elise Jennings. A turbulent eddy-viscosity surrogate modeling framework for reynolds-averaged navier-stokes simulations. *Computers & Fluids*, 227:104777, 2021.

[51] Kai Cheng and Zhenzhou Lu. Structural reliability analysis based on ensemble learning of surrogate models. *Structural Safety*, 83:101905, 2020.

[52] Estelle Guyonvarch, Elham Ramin, Murat Kulahci, and Benedek Gy Plósz. icfd: interpreted computational fluid dynamics–degeneration of cfd to one-dimensional advection-dispersion models using statistical experimental design–the secondary clarifier. *Water Research*, 83:396–411, 2015.

[53] Rongyuan Jia, Huaijie Xia, Song Zhang, Weiguang Su, and Shuhui Xu. Optimal design of savonius wind turbine blade based on support vector regression surrogate model and modified flower pollination algorithm. *Energy Conversion and Management*, 270:116247, 2022.

[54] Dian-Qing Li, Dong Zheng, Zi-Jun Cao, Xiao-Song Tang, and Xiao-Hui Qi. Two-stage dimension reduction method for meta-model based slope reliability analysis in spatially variable soils. *Structural Safety*, 81:101872, 2019.

[55] Peter Moonen and J Allegrini. Employing statistical model emulation as a surrogate for cfd. *Environmental Modelling & Software*, 72:77–91, 2015.

[56] Enrique García-Macías and Filippo Ubertini. Real-time bayesian damage identification enabled by sparse pce-kriging meta-modelling for continuous shm of large-scale civil engineering structures. *Journal of Building Engineering*, 59:105004, 2022.

[57] Iliass El Mrabti, Abdelhadi El Hakimi, Abdelhamid Touache, and Abderrahim Chamat. A comparative study of surrogate models for predicting process failures during the sheet metal forming process of advanced high-strength steel. *The International Journal of Advanced Manufacturing Technology*, 121(1-2):199–214, 2022.

[58] Olumayowa T Kajero, Rex B Thorpe, Tao Chen, Bo Wang, and Yuan Yao. Kriging meta-model assisted calibration of computational fluid dynamics models. *AIChE Journal*, 62(12):4308–4320, 2016.

[59] Ankush Aggarwal, Bjørn Sand Jensen, Sanjay Pant, and Chung-Hao Lee. Strain energy density as a gaussian process and its utilization in stochastic finite element analysis: Application to planar soft tissues. *Computer Methods in Applied Mechanics and Engineering*, 404:115812, 2023.

[60] Fei Ding and Ahsan Kareem. A multi-fidelity shape optimization via surrogate modeling for civil structures. *Journal of Wind Engineering and Industrial Aerodynamics*, 178:49–56, 2018.

[61] Li Cai, Lei Ren, Yongheng Wang, Wenxian Xie, Guangyu Zhu, and Hao Gao. Surrogate models based on machine learning methods for parameter estimation of left ventricular myocardium. *Royal Society open science*, 8(1):201121, 2021.

[62] Florian Sanchez, Marc Budinger, and Ion Hazyuk. Dimensional analysis

and surrogate models for the thermal modeling of multiphysics systems. *Applied Thermal Engineering*, 110:758–771, 2017.

[63] Li Xia, Zao-Jian Zou, Zi-Hao Wang, Lu Zou, and Hang Gao. Surrogate model based uncertainty quantification of cfd simulations of the viscous flow around a ship advancing in shallow water. *Ocean Engineering*, 234:109206, 2021.

[64] Fernando Tejero, David G MacManus, and Christopher Sheaf. Surrogate-based aerodynamic optimisation of compact nacelle aero-engines. *Aerospace Science and Technology*, 93:105207, 2019.

[65] T Boulkeraa, A Ghenaiet, S Mendez, and B Mohammadi. A numerical optimization chain combining computational fluid dynamics and surrogate analysis for the aerodynamic design of airfoils. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 228(11):1964–1981, 2014.

[66] Maulana Arifin, Ahmad Fudholi, Addy Wahyudie, and Damian M Vogt. Surrogate-based optimization of multiple-splitters radial compressor for solar hybrid microturbine. *Energy Conversion and Management: X*, 16:100332, 2022.

[67] Jay L Devore et al. *Probability and Statistics for Engineering and the Sciences*, volume 5. Duxbury Press Belmont, 1995.

[68] Rebecca Stern, Matthew C Tattersall, Adam D Gepner, Claudia E Korcarz, Joel Kaufman, Laura A Colangelo, Kiang Liu, and James H Stein. Sex differences in predictors of longitudinal changes in carotid artery stiffness: the multi-ethnic study of atherosclerosis. *Arteriosclerosis, thrombosis, and vascular biology*, 35(2):478–484, 2015.

[69] Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine learning*, 110(3):457–506, 2021.

[70] Ralph C Smith. *Uncertainty quantification: theory, implementation, and applications*. SIAM, 2013.

[71] Andrea Saltelli and Il'ya Meerovich Sobol'. Sensitivity analysis for nonlinear mathematical models: numerical experience. *Matematicheskoe Modelirovanie*, 7(11):16–28, 1995.

[72] Andrea Saltelli, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto, and Stefano Tarantola. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer physics communications*, 181(2):259–270, 2010.

[73] Jérôme Morio. Global and local sensitivity analysis methods for a physical system. *European journal of physics*, 32(6):1577, 2011.

[74] Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.

[75] Albert Tarantola, Bernard Valette, et al. Inverse problems= quest for information. *Journal of geophysics*, 50(1):159–170, 1982.

[76] Ruth D Etzioni and Joseph B Kadane. Bayesian statistical methods in public health and medicine. *Annual review of public health*, 16(1):23–41, 1995.

[77] Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 1995.

[78] Lennart Ljung. *Model validation and model error modeling*. Linköping University Electronic Press, 1999.

[79] Richard E Woods and Rafael C Gonzalez. Digital image processing, 2008. Accessed: 10-2022.

[80] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Nature, 2022.

[81] OpenCV. Opencv canny edge detection. `https://docs.opencv.org/3.4/dd/d1a/group__imgproc__feature.html#ga04723e007ed888ddf11d9ba04e2232de`, 1999. Accessed: 10-2022.

[82] Justin Liang. Canny edge detector. `https://justin-liang.com/tutorials/canny/`, 1999. Accessed: 10-2022.

[83] OpenCV. Miscellaneous image transformations. `https://docs.opencv.org/4.x/d7/d1b/group__imgproc__misc.html#ggaa9e58d2860d4afa658ef70a9b1115576a147222a96556ebc1d948b372bcd7ac59`, 1999. Accessed: 10-2022.

[84] J Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, 6:679–698, 1986.

[85] Ryan Walsh, Marie Soehl, Adam Rankin, Andras Lasso, and Gabor Fichtinger. Design of a tracked ultrasound calibration phantom made of lego bricks. In *Medical Imaging 2014: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 9036, pages 606–612. SPIE, 2014.

[86] Frank R Hampel, Elvezio M Ronchetti, Peter J Rousseeuw, and Werner A Stahel. *Robust statistics: the approach based on influence functions*. John Wiley & Sons, 2011.

[87] A Sedra, KC Smith, T Chan Carusone, and V Gaudet. Microelectronic circuits 8th edition. *Chapter*, 14:1235–1236, 2020.

[88] Chuong B Do. More on multivariate gaussians. *URL http://cs229. stanford. edu/section/more_ on_ gaussians. pdf.[Online*, 2008.

[89] Lloyd N Trefethen and David Bau. *Numerical linear algebra*, volume 181. Siam, 2022.

[90] Kevin P Murphy. *Probabilistic machine learning: an introduction*. MIT press, 2022.

[91] Jay L Devore. *Probability and Statistics for Engineering and the Sciences*. Cengage learning, 2011.

[92] John K Kruschke. Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(5):658–676, 2010.

[93] Yoav Benjamini. Simultaneous and selective inference: Current successes and future challenges. *Biometrical Journal*, 52(6):708–721, 2010.

[94] George EP Box, William H Hunter, Stuart Hunter, et al. *Statistics for experimenters*, volume 664. John Wiley and sons New York, 1978.

[95] Riccardo De Bin, Silke Janitza, Willi Sauerbrei, and Anne-Laure Boulesteix. Subsampling versus bootstrapping in resampling-based model selection for multivariable regression. *Biometrics*, 72(1):272–280, 2016.

[96] Scott I Vrieze. Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic). *Psychological methods*, 17(2):228, 2012.

[97] Shinto Eguchi and John Copas. Interpreting kullback–leibler divergence with the neyman–pearson lemma. *Journal of Multivariate Analysis*, 97(9):2034–2040, 2006.

[98] Alexei Botchkarev. Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. *arXiv preprint arXiv:1809.03006*, 2018.

[99] David Ruppert and David S Matteson. *Statistics and data analysis for financial engineering*, volume 13. Springer, 2011.

[100] Andreas K Maier, Christopher Syben, Bernhard Stimpel, Tobias Wurfl, Mathis Hoffmann, Frank Schebesch, Weilin Fu, Leonid Mill, Lasse Kling, and Silke Christiansen. Learning with known operators reduces maximum error bounds. *Nature machine intelligence*, 1(8):373–380, 2019.

[101] Ronald Christensen. *Analysis of variance, design, and regression: applied statistical methods*. CRC Press, 1996.

[102] W Holmes Finch, Jocelyn E Bolin, and Ken Kelley. *Multilevel modeling using R*. Crc Press, 2019.

[103] Johannes Oberpriller, Melina de Souza Leite, and Maximilian Pichler. Fixed or random? on the reliability of mixed-effects models for a small number of levels in grouping variables. *Ecology and evolution*, 12(7):e9062, 2022.

[104] Larry V Hedges. Effect sizes in nested designs. *The handbook of research synthesis and meta-analysis*, 2:399–416, 2009.

[105] Holger Schielzeth and Shinichi Nakagawa. Nested by design: model fitting and interpretation in a mixed model era. *Methods in Ecology and Evolution*, 4(1):14–24, 2013.

[106] Steve Midway. Data analysis in r, 2022.

[107] Andrew Gelman and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press, 2006.

[108] James W Hardin and Joseph M Hilbe. *Generalized estimating equations*. chapman and hall/CRC, 2002.

[109] Fridtjov Irgens. *Continuum mechanics*. Springer Science & Business Media, 2008.

[110] Philip M Gerhart, Andrew L Gerhart, and John I Hochstein. *Munson, Young and Okiishi's fundamentals of fluid mechanics*. John Wiley & Sons, 2016.

[111] W Michael Lai, David Rubin, and Erhard Krempl. *Introduction to continuum mechanics*. Butterworth-Heinemann, 2009.

[112] Cora JM Maas and Joop J Hox. The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational statistics & data analysis*, 46(3):427–440, 2004.

[113] Carl Edward Rasmussen. *Gaussian processes in machine learning*. Springer, 2003.

[114] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[115] Joaquin Quinonero-Candela, Carl Edward Rasmussen, and Christopher KI Williams. Approximation methods for gaussian process regression. *IEEE*, pages 203–223, 2007.

[116] Alex J Smola and Bernhard Schölkopf. *Learning with kernels*, volume 4. Citeseer, 1998.

[117] Kun Il Park, M Park, and James. *Fundamentals of probability and stochastic processes with applications to communications*. Springer, 2018.

[118] David Duvenaud, James Lloyd, Roger Grosse, Joshua Tenenbaum, and Ghahramani Zoubin. Structure discovery in nonparametric regression through compositional kernel search. In *International Conference on Machine Learning*, pages 1166–1174. PMLR, 2013.

[119] Neil Lawrence and Aapo Hyvärinen. Probabilistic non-linear principal component analysis with gaussian process latent variable models. *Journal of machine learning research*, 6(11), 2005.

[120] Michalis Titsias. Variational learning of inducing variables in sparse gaussian processes. In *Artificial intelligence and statistics*, pages 567–574. PMLR, 2009.

[121] Edward Snelson and Zoubin Ghahramani. Sparse gaussian processes using pseudo-inputs. *Advances in Neural Information Processing Systems*, 18:1259–1266, 2006.

[122] Shigeki Kaneko and Shinobu Yoshimura. Pod-galerkin fsi analysis for flapping motion. *Biomimetics*, 8(7):523, 2023.

[123] Ludovica Cicci, Stefania Fresca, Stefano Pagani, Andrea Manzoni, Alfio Quarteroni, et al. Projection-based reduced order models for parameterized nonlinear time-dependent problems arising in cardiac mechanics. *Mathematics in Engineering*, 5(2):1–38, 2023.

[124] Steven L Brunton and J Nathan Kutz. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2022.

[125] Andrea Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer physics communications*, 145(2):280–297, 2002.

[126] Thierry A Mara and Stefano Tarantola. Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering & System Safety*, 107:115–121, 2012.

[127] Michiel JW Jansen. Analysis of variance designs for model output. *Computer Physics Communications*, 117(1-2):35–43, 1999.

[128] Toshimitsu Homma and Andrea Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17, 1996.

[129] Il'ya Meerovich Sobol'. Global sensitivity indices for the investigation of nonlinear mathematical models. *Matematicheskoe modelirovanie*, 17(9):43–52, 2005.

[130] Thomas Bayes. An essay towards solving a problem in the doctrine of chances. *Biometrika*, 45(3-4):296–315, 1958.

[131] Dimitri Bertsekas and John N Tsitsiklis. *Introduction to probability*, volume 1. Athena Scientific, 2008.

[132] Richard McElreath. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC, 2018.

[133] Bradley Efron. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics: Methodology and distribution*, pages 569–593. Springer, 1992.

[134] Peter J Huber and Elvezio M Ronchetti. *Robust statistics*. John Wiley & Sons, 2011.

[135] Nicholas Metropolis and Stanislaw Ulam. The monte carlo method. *Journal of the American statistical association*, 44(247):335–341, 1949.

[136] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[137] W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 1970.

[138] Michael Betancourt. A conceptual introduction to hamiltonian monte carlo. *arXiv preprint arXiv:1701.02434*, 2017.

[139] Michael Betancourt and Mark Girolami. Hamiltonian monte carlo for hierarchical models. *Current trends in Bayesian methodology with applications*, 79(30):2–4, 2015.

[140] Charles J Geyer. Practical markov chain monte carlo. *Statistical science*, pages 473–483, 1992.

[141] Aleksander Sinek, Mateusz Mesek, Marek Rojczyk, Jan Juszczyk, Wojciech Adamczyk, Jacob Sturdy, Bartlomiej Melka, Adam Golda, Michal Nowok, Ziemowit Ostrowski, et al. Evaluating the precision and reproducibility of non-invasive deformation measurements in an arterial phantom. *Measurement*, 216:112904, 2023.

[142] S Islam, T Lehtimaki, M Juonala, M Kahonen, N Hutri-Kahonen, K Kainulainen, H Miettinen, L Taittonen, K Kontula, and J Viikari. Polymorphism of the angiotensin-converting enzyme (ace) and angiotesinogen (agt) genes and their associations with blood pressure and carotid artery intima media thickness among healthy finnish young adults—the cardiovascular risk in young finns study. *Atherosclerosis*, 188(2):316–322, 2006.

[143] Bryan Williams, Giuseppe Mancia, Wilko Spiering, Enrico Agabiti Rosei, Michel Azizi, Michel Burnier, Denis L Clement, Antonio Coca, Giovanni De Simone, Anna Dominiczak, et al. 2018 esc/esh guidelines for the management of arterial hypertension: The task force for the management of arterial hypertension of the european society of cardiology (esc) and the

european society of hypertension (esh). *European heart journal*, 39(33), 2018.

[144] Harvard apparatus. `https://www.harvardapparatus.com/`. Accessed: 2022-10-17.

[145] Endress hauser. `https://www.endress.com/en`. Accessed: 2022-10-17.

[146] Phantom. `https://www.phantomhighspeed.com/`. Accessed: 2022-10-17.

[147] Irix. `https://irixlens.com/`. Accessed: 2022-10-17.

[148] KOBOL. Kobold manual. `https://www.kobold.com/uploads/files/s4gbm_drs.pdf#:~:text=to%20Request%20Quote-,The%20Compact%20Transducer%20Amplifier%20(CTA)%20is%20a%20DC%20bridge-,transducer%20must%20be%20ordered%20together.`, March 2021.

[149] Harvard Apparatus. Ctas manual, March 2022.

[150] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[151] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[152] Wes McKinney et al. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX, 2010.

[153] Charles R. Harris, K. Jarrod Millman, and Stefan J.van der Walt. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[154] Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. Nih image to imagej: 25 years of image analysis. *Nature methods*, 9(7):671–675, 2012.

[155] Michail Vlachos, Philip Yu, and Vittorio Castelli. On periodicity detection and structural periodic similarity. In *Proceedings of the 2005 SIAM international conference on data mining*, pages 449–460. SIAM, 2005.

[156] Douglas Bates, Martin Machler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

[157] Karl-Rudolf Koch. *Parameter estimation and hypothesis testing in linear models*. Springer Science & Business Media, 1999.

[158] Bernard G Francq, Dan Lin, and Walter Hoyer. Confidence, prediction, and tolerance in linear mixed models. *Statistics in medicine*, 38(30):5603–5622, 2019.

[159] Natalie Commeau, Eric Parent, Marie-Laure Delignette-Muller, and Marie Cornu. Fitting a lognormal distribution to enumeration and absence/presence data. *International journal of food microbiology*, 155(3):146–152, 2012.

[160] Adam Loy, Heike Hofmann, and Dianne Cook. Model choice and diagnostics for linear mixed-effects models using statistics on street corners. *Journal of Computational and Graphical Statistics*, 26(3):478–492, 2017.

[161] Douglas A Reynolds. Gaussian mixture models. *Encyclopedia of biometrics*, 741(659-663), 2009.

[162] Sandip Mazumder. *Numerical methods for partial differential equations: finite difference and finite volume methods*. Academic Press, 2015.

[163] Tsutomu Ishigami and Toshimitsu Homma. An importance quantification technique in uncertainty analysis for computer models. In *[1990] Proceedings. First international symposium on uncertainty modeling and analysis*, pages 398–403. IEEE, 1990.

[164] DW Holdsworth, CJD Norley, R Frayne, DA Steinman, and BK Rutt. Characterization of common carotid artery blood-flow waveforms in normal human subjects. *Physiological measurement*, 20(3):219, 1999.

[165] Luísa C Sousa, Catarina F Castro, Carlos C António, and Elsa Azevedo. Fluid-structure interaction modeling of blood flow in a non-stenosed common carotid artery bifurcation. In *Proceedings of the 7th International*

Conference on Mechanics and Materials in Design, Albufeira, Portugal, pages 11–15, 2017.

[166] Benjamin S Aribisala, Zoe Morris, Elizabeth Eadie, Avril Thomas, Alan Gow, Maria C Valdés Hernández, Natalie A Royle, Mark E Bastin, John Starr, Ian J Deary, et al. Blood pressure, internal carotid artery flow parameters, and age-related white matter hyperintensities. *Hypertension*, 63(5):1011–1018, 2014.

[167] Litao Ruan, Wei Chen, Sathanur R Srinivasan, Meihung Sun, Hongwei Wang, Ahmet Toprak, and Gerald S Berenson. Correlates of common carotid artery lumen diameter in black and white younger adults: the bogalusa heart study. *Stroke*, 40(3):702–707, 2009.

[168] Farooq A Choudhry, John T Grantham, Ansaar T Rai, and Jeffery P Hogg. Vascular geometry of the extracranial carotid arteries: an analysis of length, diameter, and tortuosity. *Journal of neurointerventional surgery*, 2015.

[169] Jean Rossario Raj, SMK Rahman, and Sneh Anand. An insight into elasticity analysis of common carotid artery using ultrasonography. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 230(8):750–760, 2016.

[170] Adam D Gepner, Robyn L McClelland, Claudia E Korcarz, Rebekah Young, Joel D Kaufman, Carol C Mitchell, and James H Stein. Carotid artery displacement and cardiovascular disease risk in the multi-ethnic study of atherosclerosis. *Vascular Medicine*, 24(5):405–413, 2019.

[171] Topi Paananen, Juho Piironen, Michael Riis Andersen, and Aki Vehtari. Variable selection for gaussian processes via sensitivity analysis of the posterior predictive distribution. In *The 22nd international conference on artificial intelligence and statistics*, pages 1743–1752. PMLR, 2019.

[172] Aurelien Geron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc.", 2022.

[173] Zdenek Kala. Benchmark of goal oriented sensitivity analysis methods using ishigami function. *International Journal of Mathematical and Computational Methods*, 3, 2018.

[174] Bruno Sudret. Polynomial chaos expansions and stochastic finite element

methods. *Risk and reliability in geotechnical engineering*, pages 265–300, 2014.

[175] Juraj Kardos, Wouter Edeling, Diana Suleimenova, Derek Groen, and Olaf Schenk. Sensitivity analysis of high-dimensional models with correlated inputs. *arXiv preprint arXiv:2306.00555*, 2023.

[176] Oriol Abril-Pla, Virgile Andreani, Colin Carroll, Larry Dong, Christopher J Fonnesbeck, Maxim Kochurov, Ravin Kumar, Junpeng Lao, Christian C Luhmann, Osvaldo A Martin, et al. Pymc: a modern, and comprehensive probabilistic programming framework in python. *PeerJ Computer Science*, 9:e1516, 2023.

[177] Ravin Kumar, Colin Carroll, Ari Hartikainen, and Osvaldo Antonio Martín. Arviz a unified library for exploratory analysis of bayesian models in python. *Journal of Open Source Software*, 2019.

[178] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved r-hat for assessing convergence of mcmc (with discussion). *Bayesian analysis*, 16(2):667–718, 2021.

[179] Ess guidelines. `https://beast.community/ess_tutorial`. Accessed: 2024-03-31.

[180] Philip R Bevington and D Keith Robinson. Data reduction and error analysis. *McGraw-Hill, New York*, 2003.

[181] Ilker Etikan, Rukayya Alkassim, and Sulaiman Abubakar. Comparision of snowball sampling and sequential sampling technique. *Biometrics and Biostatistics International Journal*, 3(1):55, 2016.

[182] Peter J Baddoo, Benjamin Herrmann, Beverley J McKeon, and Steven L Brunton. Kernel learning for robust dynamic mode decomposition: linear and nonlinear disambiguation optimization. *Proceedings of the Royal Society A*, 478(2260):20210830, 2022.

[183] Vidhi Lalchand, Aditya Ravuri, and Neil D Lawrence. Generalised gplvm with stochastic variational inference. In *International Conference on Artificial Intelligence and Statistics*, pages 7841–7864. PMLR, 2022.

[184] Tong Teng, Jie Chen, Yehong Zhang, and Bryan Kian Hsiang Low. Scalable variational bayesian kernel selection for sparse gaussian process re-

gression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 5997–6004, 2020.

[185] Linda SL Tan. Model reparametrization for improving variational inference. *arXiv preprint arXiv:1805.07267*, 13, 2018.

Abstract

Integration of Statistical Data Analysis and Surrogate Modeling for Uncertainty Quantification, Sensitivity Analysis and Inverse Problems involving Fluid-Structure Interaction models

Local arterial stiffness is a useful marker for early detection of cardiovascular diseases. It is possible to estimate it inversely using non-invasive methods basing on measured arterial displacements. However, before implementing such a methodology clinically, it should be validated and tested in laboratory conditions. To achieve this, a testing rig for measuring arterial displacements was developed, whose reproducibility was assessed using Linear Mixed Effects Models. Basing on the test data, a Fluid-Structure Interaction model was developed, which was meant to be used in the inverse task. Ultimately, its computational overhead was deemed too demanding and a Sparse Gaussian Process Regression-based surrogate model was developed to allow for inverse estimation. The model's performance was tested and it was subsequently used in Uncertainty Quantification and Sensitivity Analysis of the original FSI model. The purpose was twofold: to gain information on the uncertainty present in the FSI model due to input data and to estimate which factors were the most influential. This was achieved making it possible to simplify the surrogate model even further and apply it in inverse estimation of the arterial stiffness based on laboratory phantom displacement data. Resulting arterial stiffness estimates proved to be very accurate. This research demonstrated the potential to apply the whole methodology to human subject test data and possible subsequent implementation of the methodology in clinical diagnostics.

Abstrakt

Integration of Statistical Data Analysis and Surrogate Modeling for Uncertainty Quantification, Sensitivity Analysis and Inverse Problems involving Fluid-Structure Interaction models

Lokalna sztywność tętcnicza jest przydatnym znacznikiem używanym we wczesnej diagnostyce chorób układu krwionośnego. Możliwym jest jej odwrotne wyznaczenie z użyciem bezinwazyjnych metod w oparciu o zmierzone przemieszczenia tętnicze. Jednakże zanim takowa metodologia zostanie zaimplementowana na poziomie klinicznym, istotnym jest ją przetestować w warunkach laboratoryjnych. W tym celu powstało stanowisko pomiarowe mierzące odkształcenia fantomów tętniczych, których powtarzalność została przeanalizowana z użyciem narzędzi statystycznych. Na podstawie tych danych został skonstruowany model Fluid-Structure Interaction, który miał zostać wykorzystany w problemie odwrotnym. Jednakże jego kosztowność pod względem czasu obliczeniowego okazała się ograniczająca i zdecydowano się wytrenować model zastępczy z użyciem techniki Sparse Gaussian Process Regression. Taki model zastępczy został następnie zastosowany do Kwantyfikacji Niepewności i Analizy Wrażliwości pełnego modelu FSI w celu eksploracji obecnej w nim niepewności oraz wykryciu najbardziej wpływowych parametrów. Informacje te posłużyły do dalszego uproszczenia modelu zastępczego, który następnie został zaimplementowany w problemie odrwotnym. Wynikająca z tego estymacją sztywności w oparciu o odkształcenia fantomów tętniczych przyniosła obiecujące wyniki, umożliwiające rozważanie dalszego testowania metodogolii na ludziach oraz potencjalnego zastosowania jej w praktyce diagnostycznej.