

MOX-Report No. 95/2024

**Neural networks based surrogate modeling for efficient uncertainty
quantification and calibration of MEMS accelerometers**

Zacchei, F.; Rizzini, F.; Gattere, G.; Frangi, A.; Manzoni, A.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it

<https://mox.polimi.it>

Neural networks based surrogate modeling for efficient uncertainty quantification and calibration of MEMS accelerometers

Filippo Zacchei^a, Francesco Rizzini^b, Gabriele Gattere^b, Attilio Frangi^c, Andrea Manzoni^a

^a*Politecnico di Milano, MOX - Dept. of Mathematics, p.za Leonardo da Vinci, 32, Milano, 20133, Italy*

^b*STMicroelectronics, Analog, Power, MEMS and Sensors Group R&D and Strategy, Cornaredo, 20007, Italy*

^c*Politecnico di Milano, Dept. of Civil and Environmental Engineering, p.za Leonardo da Vinci, 32, Milano, 20133, Italy*

Abstract

This paper addresses the computational challenges inherent in the stochastic characterization and uncertainty quantification of Micro-Electro-Mechanical Systems (MEMS) capacitive accelerometers. Traditional methods, such as Markov Chain Monte Carlo (MCMC) algorithms, are often constrained by the computational intensity required for high-fidelity (e.g., finite element) simulations. To overcome these limitations, we propose to use supervised learning-based surrogate models, specifically artificial neural networks, to effectively approximate the response of MEMS capacitive accelerometers. Our approach involves training the surrogate models with data derived from initial high-fidelity finite element analyses (FEA), providing rich datasets to be generated in an offline phase. The surrogate models replicate the FEA accuracy in predicting the behavior of the accelerometer under a wide range of fabrication parameters, thereby reducing the online computational cost without compromising accuracy. This enables extensive and efficient stochastic analyses of complex MEMS devices, offering a flexible framework for their characterization. A key application of our framework is demonstrated in estimating the sensitivity of an accelerometer, accounting for unknown mechanical offsets, over-etching, and thickness variations. We employ an MCMC approach to estimate the posterior distribution of the device's unknown fabrication parameters, informed by its response to transient voltage signals. The integration of surrogate models for mapping fabrication parameters to device responses, and subsequently to sensitivity measures, greatly enhances both backward and forward uncertainty quantification, yielding accurate results while significantly improving the efficiency and effectiveness of the characterization process. This process allows for the reconstruction of device sensitivity using only voltage signals, without the need for direct mechanical acceleration stimuli.

Keywords: Deep Learning, Neural Networks, Uncertainty Quantification, Bayesian Inverse Problems, MEMS, Computational Mechanics

1. Introduction

Micro-Electro-Mechanical Systems (MEMS) [1, 2] have emerged as one of the most prevalent technologies in today's world, manifesting in various forms across numerous devices we use daily. These systems integrate micro-structural and electronic components to miniaturize a range of functions, such as accelerometers, gyroscopes, magnetometers, and pressure sensors, down to the micron scale. However, the small scale characteristic of MEMS leads to complex fabrication processes that are influenced by multiple sources of uncertainty leading, for instance, to geometric discrepancies with respect to the nominal layout prescribed before fabrication. Fabrication variances can result in up to a 10% deviation from the intended design in standard MEMS [3, 4], significantly modifying their operational characteristics and thus requiring a comprehensive calibration process to ensure accurate measurements [5, 6, 7, 8]. Over the past few decades, continuous advancements in computational power and the development of algorithms have established numerical simulation as a cornerstone in engineering and applied sciences [9]. This is particularly relevant for solving complex differential problems, such as those governing the dynamics of MEMS [1]. As a result, it has become feasible to explore the impact of uncertainties on various design parameters [10, 11].

However, the substantial computational resources required by high-fidelity models often limit their practical utility, especially for tasks like Uncertainty Quantification (UQ), which are inherently multi-query – that is, they require repeated queries to a numerical solver. Consequently, lower-fidelity surrogate models have been increasingly adopted to enhance computational efficiency [12, 13]. While less resource-intensive, these models still provide an acceptable level of accuracy and expedite repeated computations. Surrogate models can be categorized into various types [14]: data-fit models (e.g., Kriging models [15, 16, 3, 17]), hierarchical models (e.g., simplified physics models or coarse discretizations [18]), and reduced-order models [19, 20, 21, 22, 23, 24, 25, 26, 27]. In the domain of structural mechanics, the most widely adopted techniques currently rely on reduced-order models that involve the use of modal decompositions or reduced bases [21]. Modes’ selection can be based on physical insights or system identification procedures [22], such as Proper Orthogonal Decomposition [20, 24, 28, 29, 30]. Alternatively, approaches are formulated within the framework of Nonlinear Normal Modes [31, 32, 23, 33], with several developments that have generalized the method to arbitrary order expansions and can be directly applied to finite element models, using the so-called direct parametrization of invariant manifolds (DPIM) [34, 35]. However, in recent years, a plethora of approaches for constructing data-driven surrogate models [36] have been proposed, often relying on supervised learning [15], in computational mechanics. Several methodologies have successfully integrated data-driven deep learning techniques [25] for parameter estimation and uncertainty quantification across various domains, including cardiac mechanics [37], solid mechanics [38], structural health monitoring [39], and aero-elastic systems [40], among others. In the field of MEMS devices and uncertainty quantification, the core challenge lies in determining the relationship between selected device features and the system’s behavior. Recent results have shown the feasibility of the use of deep learning-based reduced order models [41, 42] in order to simulate the mechanical behavior of MEMS also involving non-linearities [43], ultimately comparing the DPIM method with more recent model discovery strategies based on the sparse identification of nonlinear dynamics [44, 45].

In a production environment for MEMS, various tests are usually carried out to assess each device’s response and performance under external stimuli [46], also aiming at real-time auto-calibration [6]. However, inferring device characteristics directly from collected data relying on model-based methodologies often yields ill-posed problems, whose solution is even more difficult due to the unavoidable measurements’ noise. Consequently, a probabilistic framework, enabled by Bayesian model updating, emerges as a natural solution to address these issues. Bayesian model updating, in particular, provides a set of general and robust strategies for parameter estimation when dealing with physics-based models. Nonetheless, the complexity of evaluating posterior probability density functions usually calls into play computationally demanding Markov Chain Monte Carlo (MCMC) methods [47, 48, 49]. These methods require simulating device responses across different parameter sets through several (thousands of) iterations for each new device. To mitigate the computational burden, the integration of data-driven surrogate models proves highly advantageous, substantially reducing the cost per iteration. Alternative strategies to lower the computational cost include decreasing the number of required iterations, as seen in transitional MCMC [50], accelerating the convergence of the chain, e.g., through Hamiltonian MCMC [51] or the No-U-Turn sampler [52], or employing multi-level simulation fidelities, like in the case of Multi-Level MCMC methods [53]. The usage of a probabilistic learning framework for quantifying model-form uncertainties in MEMS has been developed in [54]. An approach based on Bayesian analysis for sensor calibration has been proposed in [55]. Finally, the combination of deep learning techniques with Bayesian model update has been deeply investigated in [56] or [57]. The employment of neural networks, specifically Bayesian neural networks, for the purpose of uncertainty quantification has been explored in [58].

Focusing on a real-life practical case in microsystems’ design, this work aims to characterize and identify a series of geometrical uncertainties inherent in MEMS axial capacitive accelerometers. The investigation begins with the design of a numerical test, which involves applying an active voltage to a portion of the accelerometer’s electrodes and determining the device’s response based on its characteristics. A comprehensive electro-mechanical model of the accelerometer is developed, and the finite element (FE) method is employed to generate a robust dataset of simulations. These simulations serve as the basis for training a supervised learning surrogate model, exploiting artificial neural networks (ANNs) [59], to learn the map between input

features and the device electro-mechanical response to the stimulus. Subsequently, a Bayesian procedure relying on MCMC methods is employed to infer the posterior distribution of the device’s input features from noisy data. A second configuration is then subsequently implemented to construct another surrogate model, this time focusing on predicting the device’s sensitivity, here intended as the device electrical response to a 1g acceleration, based on its features. This approach enables uncertainty quantification of the device’s sensitivity, grounded in the posterior distribution of its features. The novelty of this work lies in the application of the proposed methodology, meticulously adapted to MEMS accelerometers, and the specifically designed numerical tests to assess device characteristics. Bayesian analysis for MEMS calibration is indeed still an area to be greatly investigated [7]. The approach outlined here paves the way for integrating data-driven deep learning models with Bayesian model updating in the physical characterization of devices. Employing a data-driven mapping from parameters to responses, developed within a supervised learning framework, facilitates a significant acceleration in the evolution of the MCMC algorithm. This enhancement is attributed to the immediate assessment of likelihood for each newly generated sample. In contrast, utilizing FE simulations based on the Euler-Bernoulli approximation [60] would necessitate minutes per sample, while a comprehensive FE simulation could extend this requirement to hours for each instance. In addition the accuracy characterizing the developed surrogate model, as assessed in the work, allows to recover the same performance of the high fidelity models. Eventually, the two-stage procedure, based on the generation of two distinct surrogate models, allows us to easily forward propagate uncertainties on additional quantities of interests based on the geometrical parameters of the device. The proposed methodology is broadly applicable across a wide range of MEMS devices. However, to simplify result analysis, our study is restricted to the simpler case, yet of significant industrial relevance, of one-dimensional accelerometers.

The paper is organized as follows. Section 2 introduces one-dimensional capacitive accelerometers, highlighting their functioning and the role of fabrication inaccuracies. Section 3 details the methodology employed in this study, starting from the formulation of the model and elaborating on the development of the surrogate model and Bayesian model updating. Section 4 presents the results of our numerical experiments. After introducing a full-order model (FOM) for the numerical simulation of a class of devices, we use it to assess the devices’ behavior under varying unknown features, and to construct a comprehensive numerical dataset. Then, we discuss the training of the surrogate models and their accuracy. The identification results, obtained from noisy data using Markov Chain Monte Carlo (MCMC) methods, and the forward reconstruction of the device’s sensitivity are finally illustrated. Section 5 provides a summary of the work, some concluding remarks, and further perspectives on possible future developments.

2. Capacitive accelerometer: problem description

This section briefly outlines the operational principle of a MEMS capacitive accelerometer and examines the impact of geometric uncertainties on its performance. Capacitive sensing is the predominant mechanism in contemporary micro-accelerometers. The geometry of the x-axis MEMS accelerometer is presented in Figure 1. The device comprises a movable mass, shown in orange with no pattern, which functions as a rotor and is anchored to the substrate by two supports, depicted in blue with a chequered pattern. The device’s body is connected to the anchors through springs composed of folded beams, allowing for motion and compliance. The mass is maintained at a ground voltage. Parallel to the plates of the movable mass are placed two sets of electrodes acting as a stator and fixed to the electrode layer of the MEMS. These are denoted as the Left (in red and with a vertical stripes pattern) and Right (in gray and with a squared grid pattern) electrode groups, with an arbitrary voltage V_l and V_r . The capacitances of these groups, denoted as C_l and C_r respectively, vary over time in response to the movement of the accelerometer’s body.

With the sole purpose of elucidating the operational mechanism of this type of device, we introduce a simplified model of a 1D capacitive x-accelerometer, as depicted in Figure 2. This model features a proof mass. An electrode is attached to this proof mass, positioned equidistantly, at a distance d , from two conductive plates. These plates, referred to as the left and right electrodes, have an identical area A and are maintained at potentials V_l and V_r , respectively. If the gap d is much smaller than the dimensions of the plates, their capacitance may be approximated with the analytical formula $C = \epsilon_0 \epsilon_r A/d$. Any displacement

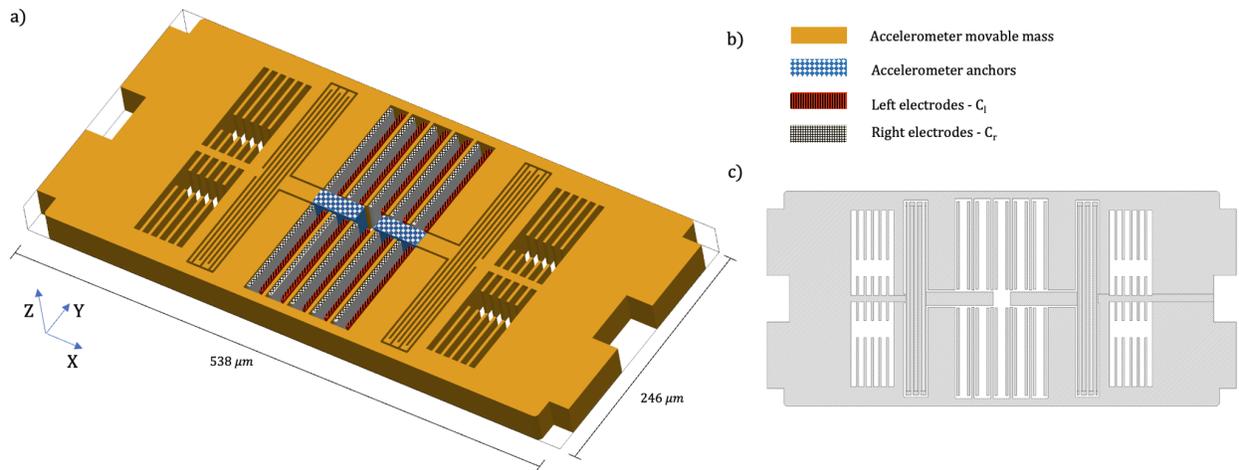


Figure 1: X-axis MEMS accelerometer: (a) Device geometry; (b) Color-coded and pattern-coded representation of device regions; (c) Device top view.

u of the proof mass along the gap direction induces differential changes in capacitance. To measure the capacitance change, in our numerical setup we will measure the charge on the electrodes and derive the value of the capacitance given the prescribed voltage on the electrodes. This can be achieved experimentally by introducing an additional readout circuit in the device. If C_l and C_r denote the capacitance of the left and right capacitors, respectively, and u is assumed positive to the right, one has:

$$\Delta C_r = \epsilon_r \epsilon_0 A \left[\frac{1}{d-u} - \frac{1}{d} \right], \quad \Delta C_l = \epsilon_r \epsilon_0 A \left[\frac{1}{d+u} - \frac{1}{d} \right]. \quad (1)$$

It is worth stressing that these formulas are based on the assumption of infinitely large plates, an approximation that is generally justified when the plate gap is significantly smaller than their area. For more accurate analytical capacitance estimations, fringe field effects, as described for instance by Palmer's formula [61], must be considered. In general, however, to calculate the capacitance variation for general devices and arbitrary configurations, numerical coupled electro-mechanical simulations are required, solving for both the displacement and electric field, and this approach will be adopted and discussed in Section 3.2. The hardest obstacles arising in the three dimensional case concern the coupled evaluation of the deformation of the geometry and the computation electric field when the analytical approximation for the capacitance is not accurate enough.

Once the capacitance variation is known, the sensitivity S of the accelerometer can be computed, defined as the ratio between the capacitance change ΔC and the external acceleration a_{0x} :

$$S := \frac{\Delta C}{a_{0x}} = \frac{\Delta C_r - \Delta C_l}{a_{0x}} \quad \left[\frac{\text{F}}{\text{g}} \right] \quad (2)$$

where formula (2) holds if the response of the device is quasi-static, i.e. if the frequency content of the external stimulus is much lower than the first eigen-frequency of the accelerometer. Under these conditions, in the operating regime of the device, the accelerometer behaves linearly. The capacitance change ΔC is directly proportional to the external acceleration a_{0x} , providing a constant value for S that is independent of a_{0x} [1]. The sensitivity of a capacitive accelerometer is significantly affected by the material, structural, and geometric features of the MEMS design. Variations introduced during the fabrication process can alter the structural configuration of the accelerometer. These variations can influence both the electrode capacitance and the device's elastic stiffness, thereby affecting the accelerometer's overall sensitivity.

The final objective of our research is the development of a cost-effective testing method to be employed during the production phase of MEMS accelerometers for sensitivity calibration. Typically, sensitivity is

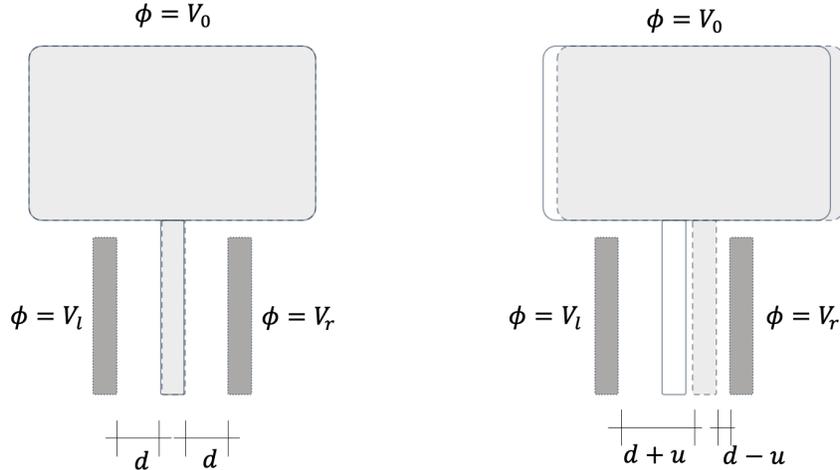


Figure 2: Parallel-plate sensing model.

determined by conducting mechanical tests, where the devices are subjected to external accelerations and their displacements are recorded. However, these tests are labor-intensive and require to move accelerometers to specialized equipment. In contrast, tests based on electrical activation can be more conveniently carried out due to easier access to electrode potentials. While mechanical testing directly yields sensitivity values, electrical activation tests can be used to assess first fabrication uncertainties which, in a second stage, can then be used to deduce the accelerometer’s sensitivity. The methodology specifically developed for addressing this issue is elaborated in the subsequent section.

2.1. The role of fabrication uncertainties

In our study, we focus our attention on three potential sources of fabrication uncertainty in the accelerometer, as illustrated in Figure 3. The first aspect deals with the mechanical offset U of the device (see Fig. 3-a). This offset can arise from process tolerances, such as uneven gaps between the differential parallel plates or mechanical stresses induced by wafer bending, which may occur due to temperature gradients during processing or operational stresses from wafer bonding. Such factors can cause the accelerometer’s proof mass to deviate from its nominal centered position, moving it closer to one set of external electrodes. This displacement can lead to a non-zero output ($\Delta C \neq 0$) even in the absence of external accelerations. In our analysis, we disregard residual stresses within the accelerometer, concentrating solely on geometrical offsets.

As a second possible fabrication defect, we consider the over-etching of the device (see Fig. 3-b). This is defined as a deviation in the in-plane geometry of the movable structure, stemming from variations in temperature and etchant concentration, as outlined in [62]. Such fluctuations result in the device’s actual geometry differing from its intended design, typically yielding a final structure marginally smaller or larger than its nominal dimensions in all directions. For the sake of simplicity, we assume a uniform overetch O on each side of the nominal geometry. This parameter not only affects the device’s capacitance by increasing the gap between electrodes, but also alters its mechanical characteristics. In particular, it reduces the width of the commonly incorporated folded beams, thereby adjusting the device’s stiffness and consequently changing its eigen-frequencies.

Lastly, we consider as unknown the device’s thickness T , defined as the height of the mechanical layer of the MEMS device and we include it as a further uncertain parameter (Fig. 3-c).

3. Methodology

In this section, we detail the methodology adopted for characterizing a MEMS accelerometer from a mechanical point of view. Our primary objective is to investigate the impact of uncertain geometric parameters

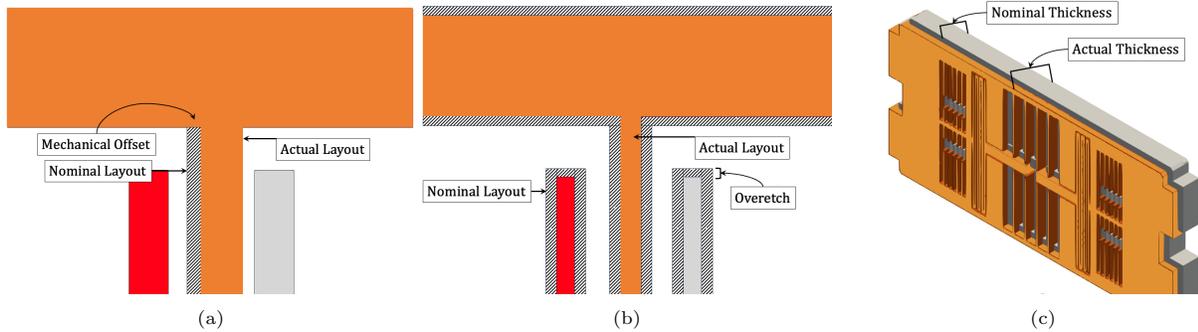


Figure 3: Fabrication uncertainties in the MEMS accelerometer: (a) Mechanical offset; (b) Over-etching; (c) Thickness variability.

on the device’s performance relying on the solution of a parameter estimation problem, enhanced by the use of neural network (NN) based surrogate models.

To build a comprehensive dataset for training our surrogate models, we performed a series of finite element analysis (FEA) simulations. These simulations were controlled by two independent input parameters: the electric voltage of the electrodes (which can differ between the left and right electrodes) and the external acceleration. We conducted two distinct sets of simulations. In the first set, we focused on the electro-mechanical response of the device under varying electric voltage stimuli, assuming zero external acceleration. Specifically, the electric voltage on the left electrodes was maintained at the same level as the proof mass, while the voltage of the remaining electrodes varied over time. This variation induced movement in the proof mass due to the electrostatic force, leading to a dynamic change in capacitance between the left and right electrodes. This capacitance variation was recorded as the response of the device to the external electric voltage stimulus. In the second set of simulations, we fixed the electric voltage of all electrodes at an operational value of 0.8 V and applied an external acceleration a_{0x} of $1g$. We recorded the resulting capacitance variation between the left and right electrodes to assess the sensitivity of the device.

Based on these data, we developed two separate surrogate models to approximate the functional relationships between the device parameters and its electro-mechanical responses. Once trained, the first surrogate model has been employed within a Bayesian model updating procedure to infer the unknown input parameters from observed device responses in the case of active voltage stimulus. Compared to a least-square estimation, a Bayesian framework allows to treat input parameters as random variables themselves, therefore providing their posterior probability distributions informed by the data, and not only a point estimate. This is of key importance not only for the chance to derive any desired statistics from the posterior distribution (like, e.g., its mean, its covariance, the maximum a posteriori, to name a few examples) but also – even more importantly – for the possibility to sample from the posterior distribution and propagate the uncertainty from the (estimated) parameters to any output quantity of interest such as, e.g., the sensitivity of the device, in the spirit of forward uncertainty quantification. This latter task requires to (i) sample from the posterior distribution and (ii) exploit the second surrogate model – once trained – for the sake of output evaluation. Overall, the very rapid evaluation capability of both surrogate models significantly reduces the computational time required for each sample generation from the posterior distribution, and the subsequent forward UQ analysis. This efficiency is crucial, considering the potentially high number of iterations needed to accurately capture the posterior distribution’s characteristics, and the need to quantify uncertainty in (possibly, several) output quantities of interest. A possible alternative is to assume a fixed shape for the posterior distribution, which leads to a variational approach typically relying on a multivariate Gaussian distribution. This generally simplifies the inverse problem and accelerates the resolution. However, thanks to the efficiency provided by the surrogate model, we can rely on MCMC sampling at similar costs. This method can characterize a posterior distribution of any shape, providing additional insights. A schematic description of the methodology is shown in Fig. 4.

We observe that an alternative methodology might involve the direct formulation of a surrogate model

connecting the device’s sensitivity to electro-mechanical response to the active voltage. However, due to the presence of multiple and distinct combinations of geometric parameters that yield identical sensitivities but different responses to voltage stimuli, a direct mapping from sensitivity to voltage response is impracticable. Consequently, the construction of a singular-stage surrogate model, along with its associated identification routine, is unviable. The two-stage approach we propose is therefore favoured not only for its operational flexibility but also for its congruence with the simulation workflow.

3.1. Problem formulation

We start by presenting an overview and establish the formal notation for the modeling principles and assumptions that constitute the high-fidelity model used throughout our work. Our focus is the electro-mechanical problem associated to a MEMS device under external stimuli. We start by considering a device occupying a spatial volume, denoted as $\Omega_0 \subset \mathbb{R}^3$, at a null electric potential, specifically $V_0 = 0$ [V]. The device’s motion is analyzed in a non-inertial reference frame attached to the MEMS system. The spatial domain in the reference configuration, represented by the material coordinates \mathbf{X} , undergoes deformation characterized by the displacement field $\mathbf{x} = \boldsymbol{\chi}(\mathbf{X}) = \mathbf{X} + \mathbf{u}(\mathbf{X}, t)$ during the time window $t \in \mathcal{T} = (0, t_{\text{final}})$. This deformation results in an actual configuration denoted by Ω , see Figure 5, which shows a schematic representation of the problem in a simplified geometry. The boundary of the reference configuration, $\partial\Omega_0$, is subdivided into two components: the Dirichlet boundary $\partial\Omega_{0D}$ and the Neumann boundary $\partial\Omega_{0N}$. Dirichlet boundary conditions are characterized by the imposition of a null displacement, while Neumann conditions are associated with the application of electric forces. Hereon, we denote by \mathbf{N} and \mathbf{n} the normal vector to the device surface in the material and the current configuration, respectively. Electrostatic potentials are applied to the external electrodes of the device, thus making it subject to an electric field. The device is enclosed within an unbounded volume Ω_∞ , where Gauss’s law is prescribed, and a zero electric flux condition is assumed on its surface S_∞ . The governing equations of the system are expressed as follows:

$$\rho_0 \ddot{\mathbf{u}}(\mathbf{X}, t; \boldsymbol{\mu}) + \mathbf{C} \dot{\mathbf{u}}(\mathbf{X}, t; \boldsymbol{\mu}) - \nabla_{\mathbf{X}} \cdot \mathbf{P}(\mathbf{u}(\mathbf{X}, t; \boldsymbol{\mu}); \boldsymbol{\mu}) = -\rho_0 \mathbf{a}_0 \quad \text{in } \Omega_0 \times \mathcal{T}, \quad (3)$$

$$\mathbf{P}(\mathbf{u}(\mathbf{X}, t; \boldsymbol{\mu}); \boldsymbol{\mu}) \cdot \mathbf{N}(\mathbf{X}) = \mathbf{f}_{elec}(\mathbf{X}) \quad \text{on } \partial\Omega_{0N} \times \mathcal{T}, \quad (4)$$

$$\mathbf{u}(\mathbf{X}, t) = \mathbf{0} \quad \text{on } \partial\Omega_{0D} \times \mathcal{T}, \quad (5)$$

$$\mathbf{u}(\mathbf{X}, 0) = \mathbf{0} \quad \text{in } \Omega_0, \quad (6)$$

$$\dot{\mathbf{u}}(\mathbf{X}, 0) = \mathbf{0} \quad \text{in } \Omega_0, \quad (7)$$

$$\text{div}(\text{grad}\phi(\mathbf{x})) = 0 \quad \text{in } \Omega_\infty \setminus \Omega \times \mathcal{T}, \quad (8)$$

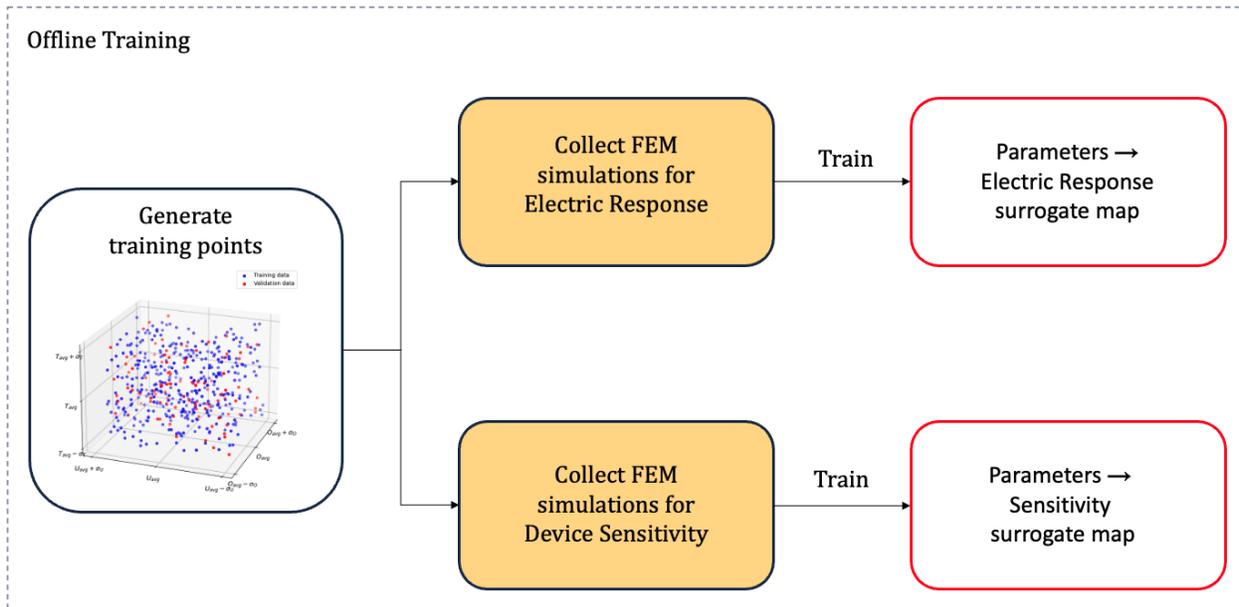
$$\phi(\mathbf{x}) = V_k(t) \quad \text{on } \partial\Omega_k, \quad (9)$$

$$\text{grad}\phi(\mathbf{x}) \cdot \mathbf{n} = 0 \quad \text{on } S_\infty. \quad (10)$$

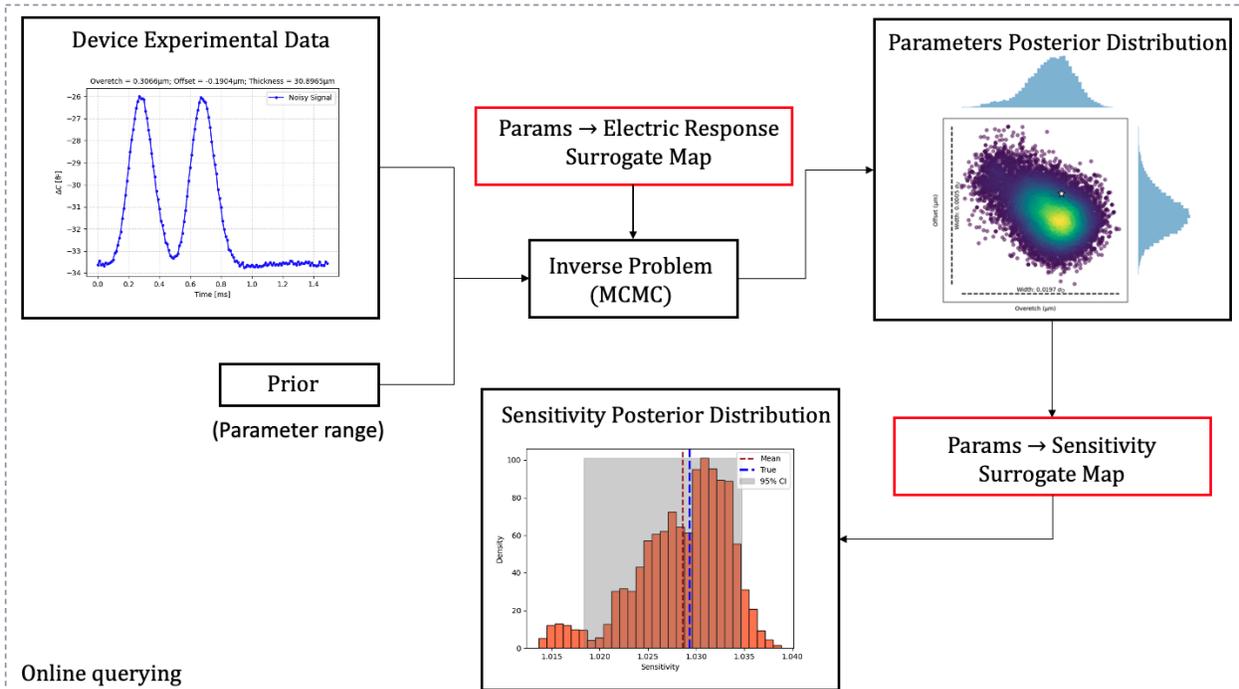
Equation (3) represents the rate of change of momentum, where ρ_0 represents the initial density in the reference configuration. The vectors $\mathbf{u}, \dot{\mathbf{u}}, \ddot{\mathbf{u}} \in \mathbb{R}^3$ denote displacement, velocity, and acceleration, respectively. The term \mathbf{a}_0 is the acceleration of the non-inertial frame of reference, as observed from an external inertial viewpoint. The terms \mathbf{C} and \mathbf{P} are the damping matrix (built according to the Rayleigh model) and the Piola-Kirchoff stress tensor, respectively. We assume mass-proportional damping for the Rayleigh model, such that $\mathbf{C} = \alpha \mathbf{M}$. The coefficient α is determined from the quality factor Q of the device. The damping factor ζ of the device is given by $\zeta = 2\omega\alpha$, where ω is the natural frequency of the device. The quality factor Q is related to the damping ratio by $Q = \zeta/\sqrt{2\pi\zeta}$. For our analysis, we set $Q = 0.55$. The parameter vector $\boldsymbol{\mu} \in \mathcal{P} \subset \mathbb{R}^p$ encapsulates the set of ~~input~~ parameters of the problem, which include all the geometrical fabrication uncertainties we want to focus on. These parameters belong to a bounded set representing the parameter space, with p denoting the number of parameters. Equations (4) and (5) represent the Neumann and Dirichlet boundary conditions, respectively.

The electrostatic pressure on the conductor surface, \mathbf{f}_{elec} , is expressed in the material configuration, originating from the following spatial formulation:

$$\tilde{\mathbf{f}}_{elec}(\mathbf{x}) = \frac{1}{2\epsilon_0\epsilon_r} \left(\epsilon_0\epsilon_r \frac{\partial\phi(\mathbf{x})}{\partial\mathbf{n}} \right)^2 \mathbf{n}, \quad (11)$$



(a) Development and training of supervised learning surrogate models.



(b) Bayesian model updating with experimental data.

Figure 4: The overall methodology of probabilistic identification framework for device geometric features and sensitivity using surrogate two-stage surrogate modelling.

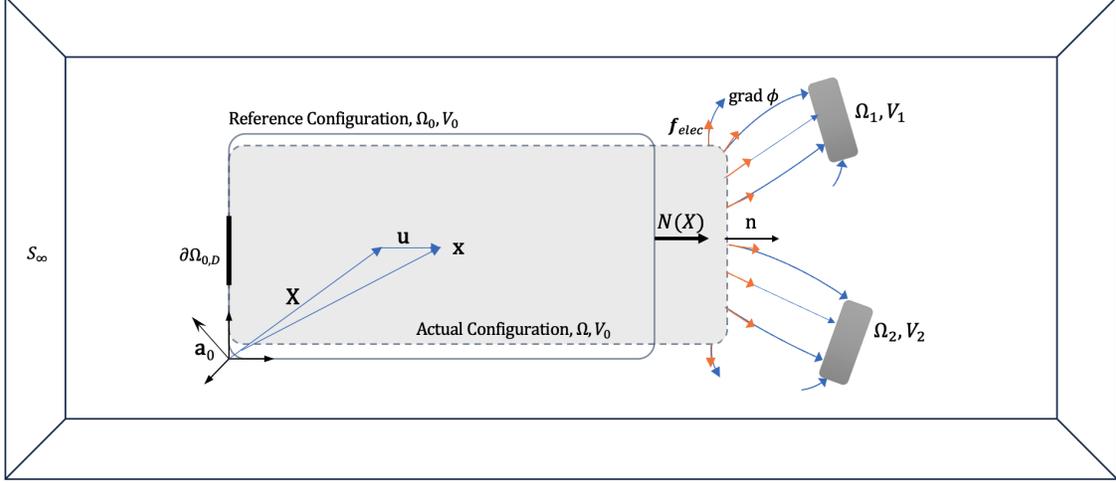


Figure 5: Schematic representation of the electro-mechanical problem coordinate scheme and reference quantities of the PDE system, in a simplified geometry.

where ϵ_0 denotes the vacuum dielectric permittivity constant and ϵ_r is the relative permittivity coefficient of air, and ϕ the electrostatic potential scalar field. The initial conditions for displacement and velocities are encapsulated in Equations (6) and (7), assumed to be zero, indicating the device's commencement from a rest position. Equation (8) represents Gauss's law for the external vacuum surrounding the device, whereas Equations (9) and (10) pertain to the active potentials applied to the fixed electrodes of the device and the condition of zero electric flux at a surface infinitely distant from the device, respectively. The variables $V_k(t)$ and \mathbf{a}_0 are treated as inputs to the model. Conversely, the variables $\mathbf{u}(\mathbf{X}, \mathbf{t})$ and $\phi(x)$ are the quantities for which we solve the equations.

In the context of our study, we adopt the hypothesis of small strains: hence displacements, along with the shape and volume variations of the solid body, are negligible compared to the body's dimensions. This hypothesis is well described by the Saint Venant-Kirchoff constitutive model:

$$\mathbf{S}(\mathbf{X}, t) = \mathcal{A}(\mathbf{X}) : \mathbf{E}(\mathbf{X}, t) \quad \text{for } (\mathbf{X}, t) \text{ in } \Omega_0 \times \mathcal{T} \quad (12)$$

where \mathbf{S} is the second Piola-Kirchoff stress, \mathbf{E} is the Green-Lagrange strain tensor, which is expressed as $\mathbf{E} = 1/2 (\nabla_{\mathbf{X}} \mathbf{u}(\mathbf{X}, t) + \nabla_{\mathbf{X}}^T \mathbf{u}(\mathbf{X}, t) + \nabla_{\mathbf{X}}^T \mathbf{u}(\mathbf{X}, t) \cdot \nabla_{\mathbf{X}} \mathbf{u}(\mathbf{X}, t))$ and \mathcal{A} is the fourth-order elasticity tensor, which, in the context of isotropic materials, is determined by specifying the values for the Young's modulus E and Poisson's ratio ν . The components of the tensor \mathcal{A} , denoted as \mathcal{A}_{ijkl} , are given by:

$$\mathcal{A}_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}),$$

where λ and μ are the Lamé parameters related to E and ν by:

$$\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)} \quad \mu = \frac{E}{2(1+\nu)}.$$

Here, δ_{ij} represents the Kronecker delta, defined as $\delta_{ij} = 1$ if $i = j$ and 0 if $i \neq j$. Consequently, the first Piola-Kirchoff stress tensor is $\mathbf{P} = (\mathbf{I} + \nabla_{\mathbf{X}} \mathbf{u}(\mathbf{X}, t)) \cdot \mathbf{S}(\mathbf{X}, t)$.

The problem described so far can be discretized (in both space and time) to obtain a high fidelity solver relying, e.g., on the FE method. It is noteworthy that our formulation utilizes material coordinates to describe mechanical equilibrium and spatial coordinates for electrostatic equilibrium, thus adding a further layer of complexity to the development of intrusive reduced order models. Despite in principle possible, the use of projection-based reduced order models (ROMs) exploiting, e.g., the proper orthogonal decomposition for the construction of a low-dimensional subspace and a (Petrov-)Galerkin projection for the generation of

the reduced-order problem, becomes rather involved in the case of coupled problems, due to the intrusive character of the method, the need to account for every variable of the problem in the ROM, and the usually complex parameter dependencies [24]. Recent strategies relying on projection-based ROMs in computational mechanics can be found, e.g., in [30, 63, 20, 64, 65, 29, 28]. For the aforementioned reasons, and given the specific need of approximating only input-output relationship in the calibration context at hand, we rather opt for data-driven surrogate models to approximate the output quantities of interest in this study. Indeed, in the current case, the output quantities of interest—specifically, the time-varying capacitance variation of the device and its sensitivity—exhibit a smooth dependency on the input parameters. This characteristic simplifies the task of approximating this relationship. Nevertheless, the availability of a set of snapshots from the high-fidelity solver, and the relatively fast training process of the surrogate models at hand, shall make the construction of further models to approximate different output quantities of interest still depending on the electro-mechanical behavior of the MEMS at hand quite straightforward.

3.2. The development of surrogate models

We rely therefore on the FE method as high-fidelity numerical solver to generate snapshots data, forecasting the response of our device under various external stimuli. To this aim, we design a numerical experiment that involves dynamically varying the active potential $V_k(t)$, following a predetermined voltage profile on one set of the device’s electrodes, while the external acceleration a_{0x} is assumed in this case to be zero. The resulting electrostatic forces prompt the accelerometer to move towards the activated electrodes even in absence of external accelerations, causing a time-varying capacitance change ΔC . This output response is influenced by the device’s uncertain parameters $\boldsymbol{\mu}$ (offset, overetch, thickness). We aim to learn a map $(\boldsymbol{\mu}, t) \mapsto \Delta C$, in order to assess this relationship efficiently, for several combinations of our input parameters. To this aim, we employ *artificial neural networks* trained using the numerical data derived from our high-fidelity simulations. Once trained, this surrogate model will then be exploited inside a Bayesian sampler for a more rapid – yet accurate – evaluation of the input-output map for any new input parameters’ value.

Simultaneously, we conduct a numerical simulation to assess the accelerometer’s sensitivity, characterized by its unknown parameters. This is done under a constant external acceleration \mathbf{a}_0 and uniform voltage across all electrode sets. Through this approach, we construct a new input-output relationship $(\boldsymbol{\mu}) \mapsto S$, for which a second surrogate model is developed. The primary application of this second surrogate is to facilitate forward uncertainty quantification of the device’s sensitivity, utilizing the posterior distribution obtained from Bayesian sampling.

3.2.1. Artificial neural networks

In this study, we focus on a NN architecture tailored for scalar map regression [59]. This architecture consists of L hidden layers, each containing N_l neurons, where the variable l specifies the layer number within the range $1 \leq l \leq L$. The network employs a nonlinear activation function ψ_l for each layer. The input vector $\mathbf{x}^{in} \in \mathbb{R}^{d_{in}}$ is processed through these layers to produce a scalar output $x^{out} \in \mathbb{R}$, prescribing $N_L = 1$. The inter-layer connectivity is defined by the weights W_{ij}^l , where j and i represent the neuron indices in the $(l-1)$ -th and l -th layers, respectively. Bias parameters are denoted as b_i^l for the i -th neuron in the l -th layer, with indices satisfying $1 \leq j \leq N_{l-1}$ and $1 \leq i \leq N_l$. The neural network’s functional mapping from \mathbf{x}^{in} to x^{out} is established through iterative linear transformations and nonlinear activation across each layer. The output of each neuron in layer l is a combination of the weighted sum of outputs from the previous layer $(l-1)$ and the bias, subsequently passed through the nonlinear activation function ψ_l . Mathematically, this

process is represented as follows:

$$x_i^0(\mathbf{x}^{in}) = x_i^{in} \quad 1 \leq i \leq d_{in} \quad (13)$$

$$x_i^l(\mathbf{x}^{in}) = \psi_l \left(b_i^l + \sum_{j=1}^{N_{n_l-1}} W_{ij}^l x_j^{l-1}(\mathbf{x}^{in}) \right) \quad 1 \leq i \leq N_{n_l}, 1 \leq l < L \quad (14)$$

$$x^{out} = b^L + \sum_{j=1}^{N_{L-1}} W_j^L x_j^{L-1}(\mathbf{x}^{in}). \quad (15)$$

For the problems at hand, we aim to approximate a multivariate scalar function, denoted by $x^{out} = f(\mathbf{x}_{in})$, using a neural network-based surrogate model, $\mathbf{f}_{NN}(\cdot; \mathbf{W}, \mathbf{b}) : \mathbb{R}^{d_{in}} \rightarrow \mathbb{R}^{d_{out}}$. This model is trained on a set of input-output pairs $\{(\mathbf{x}^{(k)}, y^{(k)})\}_{k=1}^{N_{\text{data}}}$, where \mathbf{W} and \mathbf{b} encapsulate all the network's weights and biases, respectively. The surrogate model is trained by minimizing the Mean Squared Error (MSE) loss function, augmented with L^2 regularization weighted by a regularization coefficient λ . The final loss function $\mathcal{L}(\mathbf{W}, \mathbf{b})$ to be minimized is defined as:

$$\mathcal{L}(\mathbf{W}, \mathbf{b}) = \frac{1}{N_{\text{data}}} \sum_{k=1}^{N_{\text{data}}} \left(\mathbf{f}_{NN}(\mathbf{x}^{(k)}; \mathbf{W}, \mathbf{b}) - y^{(k)} \right)^2 + \lambda \sum_{i,j} W_{ij}^2.$$

The parameter N_{data} represents the total number of pairs, derived from FE simulations. This number coincides with the product of the total simulations count N_{sim} and the number of time steps per simulation N_t for the electric response surrogate map (defined as $\mathbf{f}_{NN}^{(I)}$). Conversely, it corresponds to simply N_{sim} for the sensitivity surrogate map (defined as $\mathbf{f}_{NN}^{(II)}$).

3.3. Parameter identification and uncertainty quantification

We now introduce and formalize the application of a Bayesian model updating strategy, specifically tailored for estimating the sensitivity of MEMS accelerometers and providing uncertainty quantification. At the end of the production line of a device accelerometer, electrical inputs can be easily induced. By monitoring the electro-mechanical response of the device under electrical stimuli, we can infer the sensitivity of the device using the following approach, based on the two surrogate models previously developed. Initially, we employ a Markov Chain Monte Carlo (MCMC) analysis to derive the posterior distribution of the device parameters. This procedure can characterize the fabrication inaccuracies, starting from an experimental signal acquired in the factory line. Our analysis is based on numerical data that have been artificially corrupted with external noise to mimic the experimental data, acknowledging the inherent noise in the observed responses. The first surrogate model is used to evaluate the likelihood at each MCMC step. Subsequently, repeated queries to the second surrogate model are performed by drawing samples from the posterior distribution of the parameters, in order to recover the predictive posterior distribution of the device's sensitivity, based on its electro-mechanical response. This two-stage approach—first focusing on parameter estimation, and then on uncertainty propagation—enables a comprehensive uncertainty quantification of the MEMS device.

3.3.1. Bayesian inference

In our study, we adopt a Bayesian inference approach to infer the posterior distribution of device parameters. This process relies on the choice of a prior distribution on the input parameters, a set of noisy, observational data (of the capacitance change) and a likelihood function that assimilates those data points. The posterior distribution $\pi(\boldsymbol{\mu} | \Delta \mathbf{C})$ of the input parameters $\boldsymbol{\mu}$, given the data $\Delta \mathbf{C}$, is obtained through Bayes' theorem as

$$\pi(\boldsymbol{\mu} | \Delta \mathbf{C}) = \frac{\pi(\Delta \mathbf{C} | \boldsymbol{\mu}) \pi(\boldsymbol{\mu})}{\pi(\Delta \mathbf{C})}, \quad (16)$$

where $\Delta \mathbf{C}$ is the experimental data vector collecting the recorded measurements on the capacitance change for a tested device characterized by unknown true parameters $\boldsymbol{\mu}^{\text{true}}$, that in our case it coincides with the

vector $\Delta \mathbf{C}(\boldsymbol{\mu}^{\text{true}}) := \{\Delta C(\boldsymbol{\mu}^{\text{true}}; t)\}_{t=t_1, \dots, t_{N_t}}$. The value $\pi(\boldsymbol{\mu})$ represents the prior distribution, $\pi(\Delta \mathbf{C} | \boldsymbol{\mu})$ the likelihood function, while $\pi(\Delta \mathbf{C})$ is the marginal distribution, acting as a normalization factor.

Regarding the prior, we assume a uniform distribution on the three parameters, within a suitable bound. Based on the assumptions that measurement errors are present, and that they are independent and identically normally distributed with zero mean and fixed variance σ^2 accounting for noise in measurement acquisitions, we can approximate the relationship between input parameters and measurements as follows,

$$\Delta C(\boldsymbol{\mu}; t) = f_{NN}^{(I)}(\boldsymbol{\mu}; t) + \varepsilon$$

where $\varepsilon \sim N(0, \sigma^2)$ and $f_{NN}^{(I)}(\boldsymbol{\mu}; t)$ is the response provided by the first surrogate model at time t given $\boldsymbol{\mu}$. In this case, the likelihood function takes the following form [66]:

$$\pi(\Delta \mathbf{C} | \boldsymbol{\mu}) := \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{N_t}{2}} \exp \left[-\frac{SS_q}{2\sigma^2} \right], \quad (17)$$

where SS_q is the sum of the squares error, given by

$$SS_q := \sum_{i=1}^{N_t} \left[\Delta C(\boldsymbol{\mu}^{\text{true}}; t_i) - f_{NN}^{(I)}(\boldsymbol{\mu}; t_i) \right]^2, \quad (18)$$

and $\Delta C(\boldsymbol{\mu}^{\text{true}}; t_i)$ denotes the measured data at time step t_i .

To compute the posterior distribution $\pi(\boldsymbol{\mu} | \Delta \mathbf{C})$ without directly determining $\pi(\Delta \mathbf{C})$, we rely on a Monte Carlo Markov Chain method, specifically the Metropolis-Hastings algorithm [67]. This algorithm sequentially generates a sample series to approximate the desired distribution, with each new sample based on the previous one. At each iteration i , the algorithm proposes a candidate $\boldsymbol{\mu}'$, based on the current sample $\boldsymbol{\mu}^i$. This candidate is either accepted—forming the basis for the next iteration—or rejected, in which case $\boldsymbol{\mu}^i$ is retained. The acceptance probability is determined by the ratio:

$$\alpha = \frac{\pi(\Delta \mathbf{C} | \boldsymbol{\mu}') \pi(\boldsymbol{\mu}')}{\pi(\Delta \mathbf{C} | \boldsymbol{\mu}^i) \pi(\boldsymbol{\mu}^i)}, \quad (19)$$

comparing the proposed and current samples in relation to the desired distribution.

The posterior samples can then be used to estimate the input with the maximum posterior density and to estimate expectations of functions of interest with respect to the posterior distribution. For instance, in our case we are interested in the expected value of S over the posterior distribution, which can be approximated as:

$$\mathbb{E}[S] = \int f_{NN}^{(II)}(\boldsymbol{\mu}) \pi(\boldsymbol{\mu} | \Delta \mathbf{C}) d\boldsymbol{\mu}. \quad (20)$$

where $f_{NN}^{(II)}$ is the map from parameters to device sensitivity given by the second surrogate model.

4. Results and discussion

In this section, we detail the results and validation of our methodology on an x-axis capacitive accelerometer. The investigation is structured as follows: initially, we describe the full-order model (FOM) obtained through the FE method. Next, we detail the development and accuracy assessment of the introduced surrogate models. Lastly, we present our findings from the parameter identification process and uncertainty quantification concerning the device's sensitivity.

4.1. Problem description and FOM results

Two distinct simulation configurations are employed to characterize the device’s response to an active voltage signal and to determine its sensitivity. The first configuration represents an electro-mechanical test, which could be used at the end of the device fabrication process for product characterization. In this setup, an active voltage is applied to the right electrodes, denoted as C_r , while the left electrodes and the device body are maintained at ground potential. This active voltage induces an electrostatic force, causing the device to move towards the active electrodes, and consequently altering the values of both C_l and C_r . The active voltage profile is designed in the form of two complete sine wave periods, varying from 0 V to 1.8 V. The mathematical representation of this profile is as follows:

$$V_r = 0.9 \times (1 - \cos(2\pi ft)) \quad [\text{V}] \quad \text{for } 0 \leq t \leq 2T \quad [\text{s}], \quad (21)$$

$$V_r = 0 \quad [\text{V}] \quad \text{for } t > 2T \quad [\text{s}], \quad (22)$$

$$V_l = 0 \quad [\text{V}] \quad \text{for } t \geq 0 \quad [\text{s}], \quad (23)$$

where f is the impulse frequency set to 2500 Hz (close to resonant frequency) and $T = 1/f = 0.4$ ms. This specific profile has been chosen as a good trade-off between simplicity, as each period of the sine wave necessitates simulation time, and generated information. The choice of a two-periodic input is instrumental in enhancing the identifiability of the device’s frequency response, which is highly affected by the device overetch. Further periods would result in additional information, however augmenting the cost of each simulation and the complexity of the surrogate model. As a final output of the simulation, we retain the capacitance difference in time between right and left electrodes, i.e. $\Delta C(t) := \Delta C_r(t) - \Delta C_l(t)$, for t ranging from 0 to $1.5\text{ms} = 3.75T$, therefore leaving time to the device to come back to the rest position after the periodic impulse. A second configuration is designed to mimic operating conditions, where both electrode groups are set at 0.69 V and an external axial acceleration of $a_0 = 1g$ is applied to the MEMS accelerometer. This setup is aimed at determining the device’s sensitivity, which is defined as $S := (\Delta C_{final} - \Delta C_0)/1g \quad [F/g]$, where ΔC_0 stands for the capacitance difference before the acceleration is applied, which has been introduced with respect to (2) to account for initial asymmetries due to offset, and ΔC_{final} at the equilibrium position when $1g$ acceleration is present.

Both simulations have been carried out using Coventor MEMS+TM [68]. This software employs Euler-Bernoulli elements to model the beams and conformal mappings for describing electrostatic forces, using eight integration points per electrode. For the first configuration, we use the forward Euler scheme damping the solution every 0.01ms , for a total of $N_t := 150$ time steps. It is important to note that the predicted displacements in these simulations are minimal relative to the size of the elements. This allows us to neglect geometric non-linearities, as the nonlinear effects in this scenario are predominantly induced by electro-mechanical interactions. The use of the classical beam theory in Coventor MEMS+ accelerates the simulations by up to 100 times compared to a full FE simulation. Specifically, the speed-up is obtained since the components of the device are modeled using pre-assembled blocks characterized by predefined governing equations: the proof mass is treated as a rigid body, while the folded beams are represented using Euler-Bernoulli beam theory. This approach drastically reduces the number of degrees of freedom compared to traditional Finite Element simulations, resulting in a total number of dofs in the order of the number of distinct components of the device. Furthermore, to handle the electric field calculations efficiently, conformal mapping techniques are employed. Initially, the electrode coupling is resolved in a simple reference configuration, and then remapped for each new geometric arrangement. This approximation is reasonable in our case since the displacement is very small and the electrodes have a parallel configuration. This method simplifies the solution of the electric field problem, which is typically computationally intensive in Finite Element simulations due to the required high refinement. This acceleration is particularly advantageous in our scenarios, where the device’s behavior needs to be tested across several geometric parameters to generate an extensive training dataset. However, the simulation times are still significant, limiting the feasibility of using this software for repeated iterations, such as during the calibration stage with a MCMC sampler, thus highlighting once more the need of employing a surrogate model.

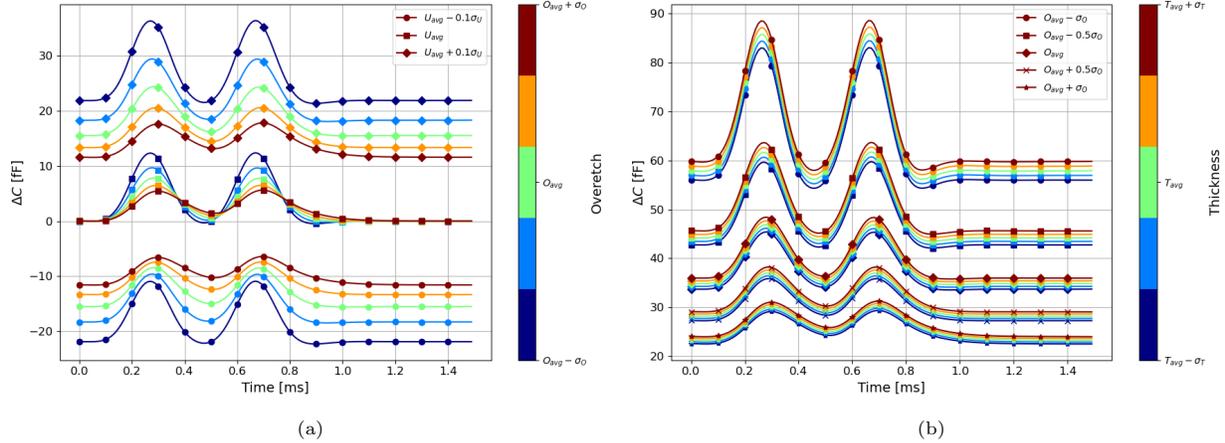


Figure 6: Numerical results obtained using Coventor MEMS+ for the first simulation configuration: (a) Fixed thickness $T = 30.0\mu\text{m}$; (b) Fixed offset $U = 0.0\mu\text{m}$.

4.1.1. Coventor results

The results illustrating the time-dependent variation of the capacitance (ΔC) for a range of thickness (T), offset (U), and overetch (O) values, are reported in Figure 6. The plot shows the influence of the offset parameter mainly translating the steady state value of ΔC , as well as the impact of the overetch on the capacitance response amplitude to the impulse and frequency. Conversely, the thickness parameter exhibits a relatively minor effect on the output within the investigated range. This may lead to challenges in its identification, particularly in scenarios with high noise levels. Snapshots of the solution for the displacement and voltage field for the second configuration of the simulation at equilibrium, considering a device with $0.5\mu\text{m}$ overetch, $-0.1\mu\text{m}$ offset and $30\mu\text{m}$ thickness, are plotted in Figure 7.

4.2. Surrogate model results

In this section, we present the numerical results obtained from the surrogate models applied to two simulation configurations. For these simulations, fully connected neural networks (FCNN) are employed to build our surrogate models. The training dataset is made by 800 selected triples in the parameter space

$$[U_{\text{avg}} - \sigma_U, U_{\text{avg}} + \sigma_U] \times [O_{\text{avg}} - \sigma_O, O_{\text{avg}} + \sigma_O] \times [T_{\text{avg}} - \sigma_T, T_{\text{avg}} + \sigma_T],$$

accounting for variations in three parameters: mechanical offset U , unilateral overetch O , and thickness T . Here we choose $U_{\text{avg}} = 0.0\mu\text{m}$, $O_{\text{avg}} = 0.0\mu\text{m}$, $T_{\text{avg}} = 30.0\mu\text{m}$. The beam width that constitutes the springs, adjusted for the mean overetch value, measures $2.1\mu\text{m}$. Sensitivity metrics are normalized against a baseline sensitivity S_{avg} , which corresponds to a device characterized by mean values of overetch, offset, and thickness. This normalization is represented by the ratio S , defined as the sensitivity of the device relative to S_{avg} . Sampling points within these ranges are generated using the Latin hypercube sampling technique. Due to confidentiality agreements with the company producing the accelerometer at hand, specific values for these parameters cannot be disclosed.

Neural networks are trained using the Adam optimization algorithm, and a total number of epochs equal to 100000 which was tuned by hand. The dataset was subdivided in a 80-20 training-validation split to facilitate hyper-parameter tuning within the training-validation framework. Batch training is employed using 40 mini-batches. The architectures of the neural networks and the L_2 regularization coefficient λ were calibrated using grid search optimization aimed at minimizing the validation loss, considering both the mean and variance. Parameters for training were selected through fine hand tuning, as they demonstrated a secondary influence on the overall model accuracy. Our implementation is based on the TensorFlow Keras library for deep learning [69]. Detailed information regarding the final structure of the employed NN architectures are reported in the following Tables 1 and 2.

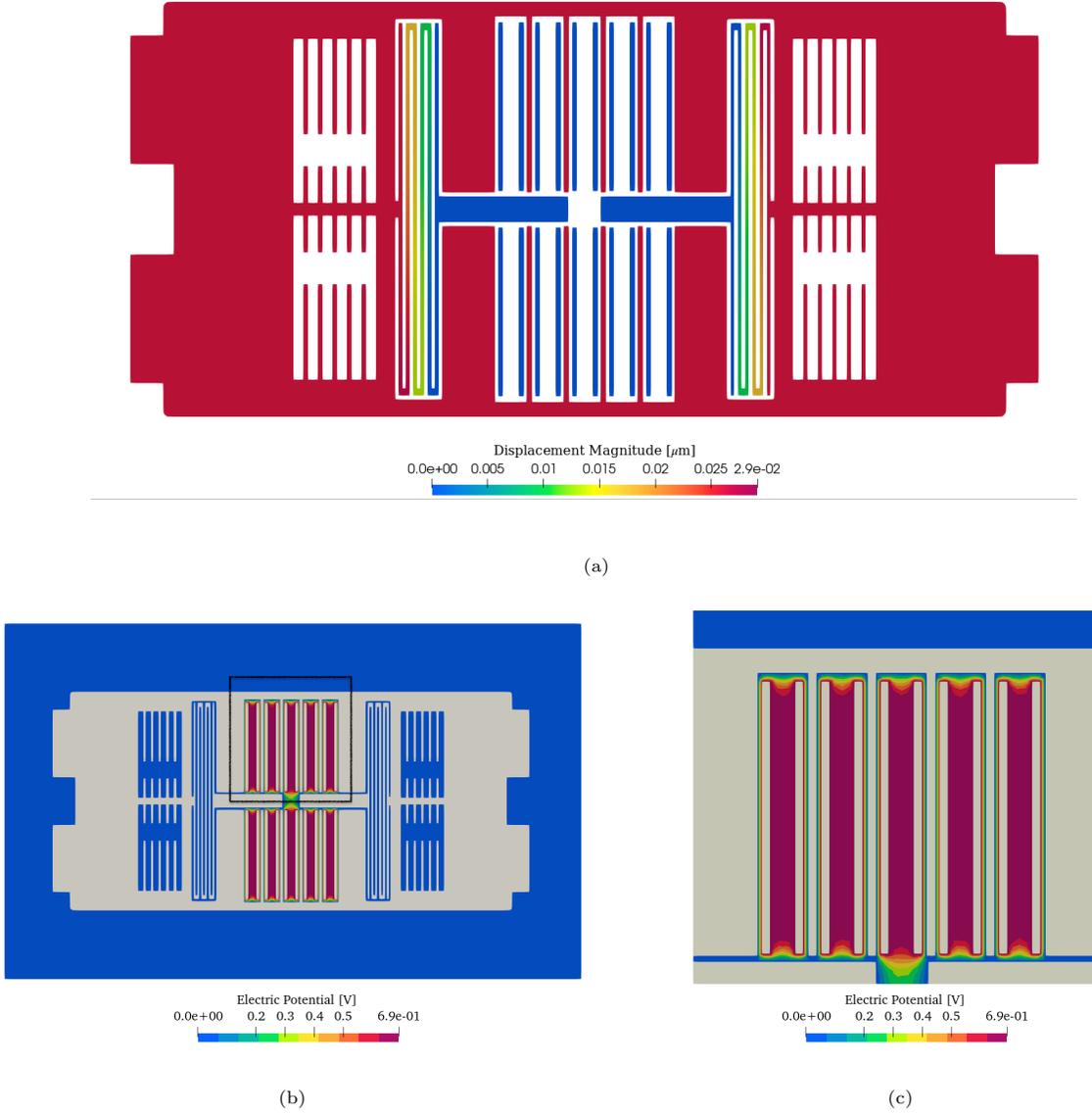


Figure 7: Snapshots of the numerical solution obtained using Coventor MEMS+ for the second simulation configuration: (a) Displacement magnitude; (b) Electric field; (c) Electric field, zoom close to the electrodes.

4.2.1. Configuration I: electro-mechanical response to an active voltage signal

In this configuration, the training dataset is made by $N_{\text{sim}} = 800$ simulation samples, generated using Coventor MEMS+. The dataset is visualized in Figure 8a. Each sample captures the time-dependent capacitance variation ΔC , considering time as an additional input parameter for the neural network. Consequently, the network is meant to provide a mapping from the geometric parameters $\boldsymbol{\mu} := \{\mu_U, \mu_O, \mu_T\}$, consisting of the values of offset, overetch and thickness, respectively, and the discrete time instant t_k , to the corresponding value of the capacitance variation ΔC :

$$f_{NN}^{(I)} : (\mu_U, \mu_O, \mu_T; t_k) \mapsto \Delta C_i(t_k) \quad \text{for } k = 1, \dots, N_t.$$

The training errors for both the training and validation datasets are depicted in Figure 8b, highlighting the absence of overfitting. Predictions from the surrogate model for two sets of input parameters outside the

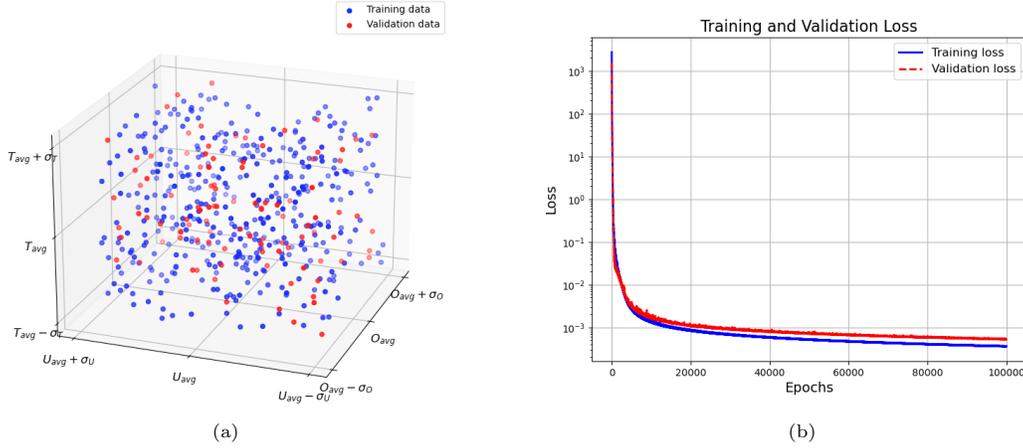


Figure 8: Training details of the first surrogate model: (a) Simulation samples, divided into training and validation dataset, used during the training process; (b) Loss function evolution during the training process considering the training and the validation dataset.

training/validation dataset are illustrated in Figure 9, demonstrating a close match with the testing data. Using an additional linearly spaced grid of samples for testing, we obtain a Pearson correlation coefficient of 99.92%. A high number of epochs and accurate fine tuning is needed in particular when very low noise signals will be the starting point of the parameter identification routine, i.e. the noise is the same order of the error of the surrogate model.

The performance of the NN models has also been compared against two different supervised learning strategies: Gaussian Process Regression (GPR) [70] and Long Short-Term Memory networks (LSTM) [71]. The former retains the same network input-output map and generally shows more resistance to overfitting, albeit at a higher computational cost due to the matrix operations involved at prediction. The latter instead leverages time-series data during predictions, implemented by integrating an LSTM cell layer within the network and inputting the complete sequence of sample time values. Performance comparisons among the three models revealed that Gaussian Process Regression (GPR) encounters challenges due to the curse of dimensionality. Achieving high accuracy requires a large number of data points, which we addressed using sparse GP regression, still leading to significant computational costs that are 50 times higher compared to NN architectures. For the problem at hand, this computational overhead makes GPR less feasible. LSTM networks and FCNN exhibit minimal performance differences, with FCNN proving advantageous due to its faster evaluation times, being approximately 10 times quicker than LSTM. Given the simple nature of the output profile, a shallow NN effectively processes the task by treating time as a direct input parameter, resulting in slightly quicker evaluation times compared to LSTM.

4.2.2. Configuration II: electro-mechanical response for an external 1g acceleration

In this configuration, we train a surrogate model using the same samples as in the previous setup, but with the objective of inferring the sensitivity of the device. We recall that sensitivities are determined by measuring the capacitance variation between the left and right electrodes under a constant external acceleration of 1g for the various combinations of device parameters. This value remains constant within the operating regime of the device, as the capacitance variation exhibits a linear relationship with the external acceleration. Therefore, once characterized, it is used in the future to infer the external acceleration starting from the recorded capacitance variation. We highlight that the mapping from parameters to sensitivity, showed in Fig. 10a, is not bijective, unlike the case in the parameter-response map to active electric voltage. This fact justifies our adoption of a two-stage approach for uncertainty quantification. Accordingly, we design a second surrogate model to represent the mapping

$$f_{NN}^{(II)} : (\mu_U, \mu_O, \mu_T) \mapsto S.$$

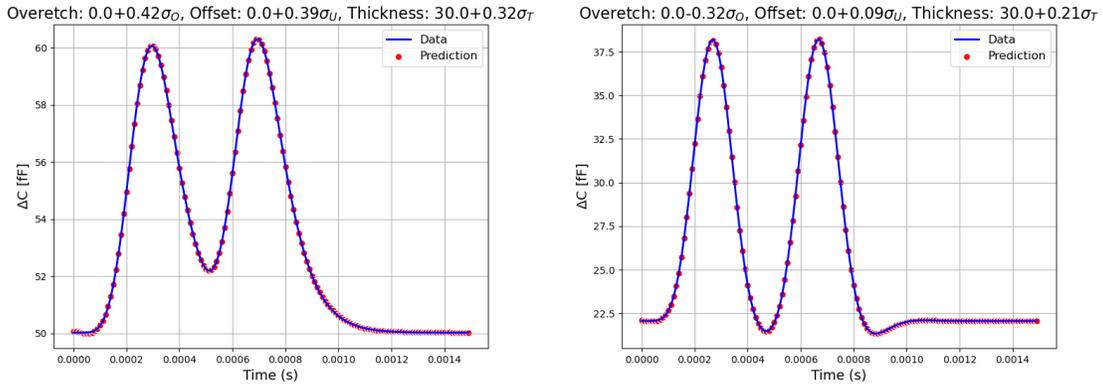


Figure 9: Comparison between testing data and surrogate model predictions for two input parameter combination unseen during training process.

The training dataset is split in the same 80-20 ratio as before. Post-training predictions, shown in Figure 10b, indicate an almost perfect alignment between the surrogate model’s predictions and the actual data. The architecture of this neural network is detailed in Table 2.

4.3. Bayesian analysis

The results obtained with our proposed methodology and the considered surrogate models are now examined within the context of an MCMC-based model updating framework. This framework is dedicated to the calibration of a one-dimensional accelerometer’s sensitivity.

The surrogate model $f_{NN}^{(I)}$ is employed to approximate the posterior probability density function $\pi(\boldsymbol{\mu} \mid \Delta\mathbf{C}_{1,\dots,N_{\text{obs}}}^{\text{exp}})$ of the geometric unknown parameters $\boldsymbol{\mu} = \{\mu_U, \mu_O, \mu_T\}$. The posterior probability is conditioned on the experimental data $\Delta\mathbf{C}_{1,\dots,N_{\text{obs}}}^{\text{exp}}$, where $N_{\text{obs}} = N_t$, representing a time series of the device’s response to an electrical signal. For experimental observations, we use testing data samples generated via Coventor MEMS+ that were not included in the training process. These data samples are numerically corrupted by an additive, independent Gaussian white noise, uncorrelated with time, with zero mean and a standard deviation defined as $\sigma := n_l S_{\text{avg}} \sqrt{B_W}$, assuming a constant noise density independent of the frequency. Here, n_l denotes the noise level (in $g/\sqrt{\text{Hz}}$), which is the noise density per frequency of the accelerometer, B_W represents the total bandwidth of the device (in Hz), and S_{avg} denotes its baseline sensitivity (in F/g). The artificial corruption introduced represents the noise that may appear due to the interference of electronic components in the experimental setup for measurements. Conservatively, we assume B_W to be 200 Hz, since the resonant frequency of the device is between 2000 and 3000 Hz, and S_{avg} to be the baseline sensitivity, in accordance with the device features. The same value of the standard deviation is used inside the likelihood, following the formula in equation (17).

Layer	Nr. of neurons	Activation Function
1	4	<i>tanh</i>
2	64	<i>tanh</i>
3	64	<i>tanh</i>
4	64	<i>tanh</i>
5	64	<i>tanh</i>
6	64	<i>tanh</i>
7	1	<i>linear</i>

Table 1: Features of dense layers in the FCNN $f_{NN}^{(I)}$. The total number of parameters is 20800, regularization coefficient is $\lambda = 10^{-12}$.

Layer	Nr. of neurons	Activation Function
1	3	<i>tanh</i>
2	32	<i>tanh</i>
3	32	<i>tanh</i>
4	32	<i>tanh</i>
5	1	<i>linear</i>

Table 2: Features of dense layers in the FCNN $f_{NN}^{(II)}$. The total number of parameters is 2176, regularization coefficient is $\lambda = 10^{-9}$.

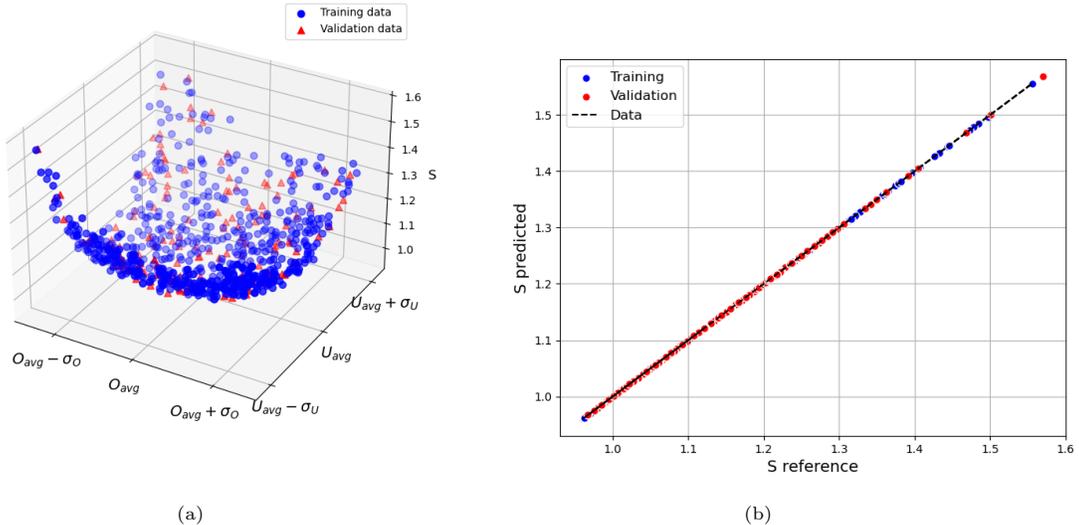


Figure 10: Details of the second surrogate model: (a) Scatter plot of sensitivity with respect to overetch and offset features, highlighting training and validation points; (b) Comparison of the predictions of the model and the reference data at the end of the training process.

Although an informative prior can offer valuable insights into the device parameters and statistical tools can be employed to design it, a uniform prior over the whole parameter space, described in section 4.2, is assumed for all the device parameters $\boldsymbol{\mu}$ to prevent potential biases in the results.

For monitoring the convergence of the Markov Chain, the Gelman and Rubin’s convergence diagnostic \hat{R} [72] is employed. This statistics is computed from n independent Markov Chains initialized at different starting points $\boldsymbol{\mu}^0$. The \hat{R} metric assesses the convergence of a multivariate chain by comparing the ratio of the between-chain variance estimate of samples to the average within-chain variance. Convergence is considered to be achieved when \hat{R} falls below a predetermined threshold, typically set at 1.01. Once the MCMC samples are obtained, we leverage the second surrogate model $f_{NN}^{(II)}$ to infer the distribution of the sensitivity, by evaluating such surrogate model for each sample of the chain.

The Bayesian procedure adopted in this study is implemented using CUQIpy [73], a comprehensive and open-source Python package designed specifically for computational UQ in inverse problems.

4.3.1. Results

In our Markov Chain Monte Carlo (MCMC) simulations, each chain starts from $N_{\text{obs}} = N_t$ observations of the time-dependent capacitance difference ΔC , associated with a specific set of input parameters $\boldsymbol{\mu} = \{\mu_U, \mu_O, \mu_T\}$. To ensure a comprehensive exploration of the parameter space, we initiate five independent random chains from distinct starting points. We start performing a least-square optimization, i.e. we minimize using gradient descent and starting from the 5 random points, the cost function $SS_q(\boldsymbol{\mu})$, as defined by equation (18). The obtained parameters $\boldsymbol{\mu}^{\text{LS}}$ are approximations of the maximum likelihood estimator and are used as the starting points of the Metropolis Hastings algorithm. The total number of samples in each chain is adjusted to achieve a Gelman-Rubin statistic \hat{R} below 1.01, resulting in a collective sample count of approximately 6,000, with a number of samples tuned by using one case as representative. To mitigate the effects of initial transients, the first sixth of the samples from each chain are discarded as part of the burn-in process. Furthermore, to reduce the auto-correlation inherent in the chains, we thin the samples by retaining only every fifth sample post burn-in. The proposal distribution within the MCMC algorithm is assumed to be a Gaussian random walk, with covariance matrix defined as the inverse of $(\mathbf{J}(\boldsymbol{\mu}^{\text{LS}})^T \mathbf{J}(\boldsymbol{\mu}^{\text{LS}}))$. Here, $\boldsymbol{\mu}^{\text{LS}}$ is the solution provided by the least square optimization, whereas \mathbf{J} is the Jacobian of the least

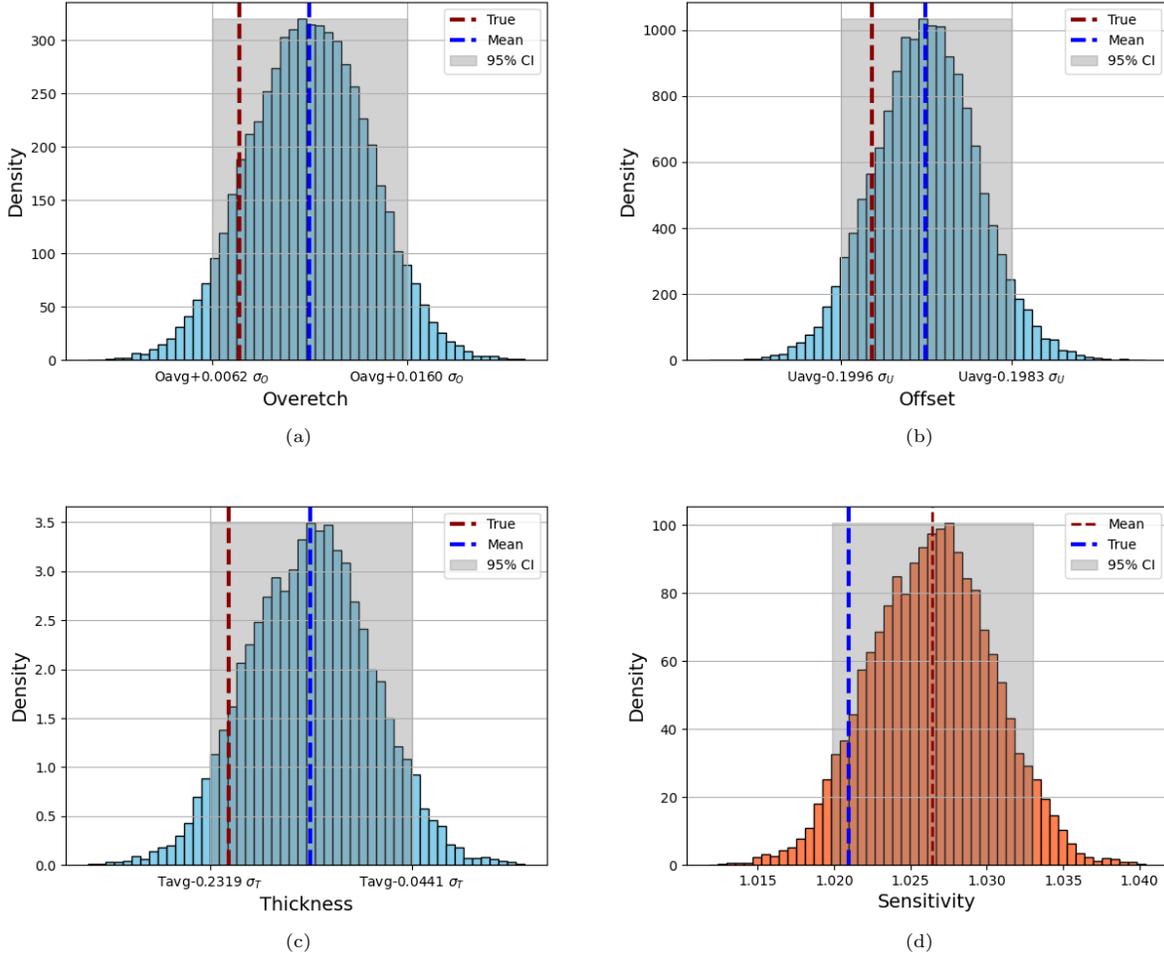


Figure 11: (a)-(c) Histogram representing densities of the posterior distribution for each input parameter obtained by the MCMC; (d) Histogram of the final density of the posterior distribution of the sensitivity of the device.

squares cost function SS_q evaluated at μ^{LS} . Additionally, the width of the random walk is dynamically adjusted throughout the simulation to maintain an acceptance rate between the optimal thresholds of 0.2 and 0.4.

Ten MCMC analyses are carried out for ten devices characterized by different geometric parameters, denoted by labels A to L, and the same noise level. The results for device A, derived under a noise factor of $n_l = 1000 \mu g / \sqrt{\text{Hz}}$, which represents a realistic value for the noise density of the device, are depicted in Figure 11. A remarkable observation from the top figures is the narrow confidence intervals for the overetch and offset parameters under this noise factor, indicating precise parameter estimation. This is particularly true for the offset, thanks to the high and distinctive effect of this parameter. In Figure 11d we report the posterior distribution obtained for the sensitivity of the same device, computed by starting from the posterior distribution of the parameters and evaluating the second surrogate model for each sample. We observe that the reconstruction error for the sensitivity is very low, and we provide again a narrow range for the confidence interval confirming the accuracy of the proposed method. We observe in addition that the distributions show negligible skewness. Indeed, the mean and the mode of each distribution are very close for each parameter and for the sensitivity, too. Results have been obtained for ten different devices, and the performances in reconstruction of the sensitivity for the remaining nine devices are reported in the Appendix. The study

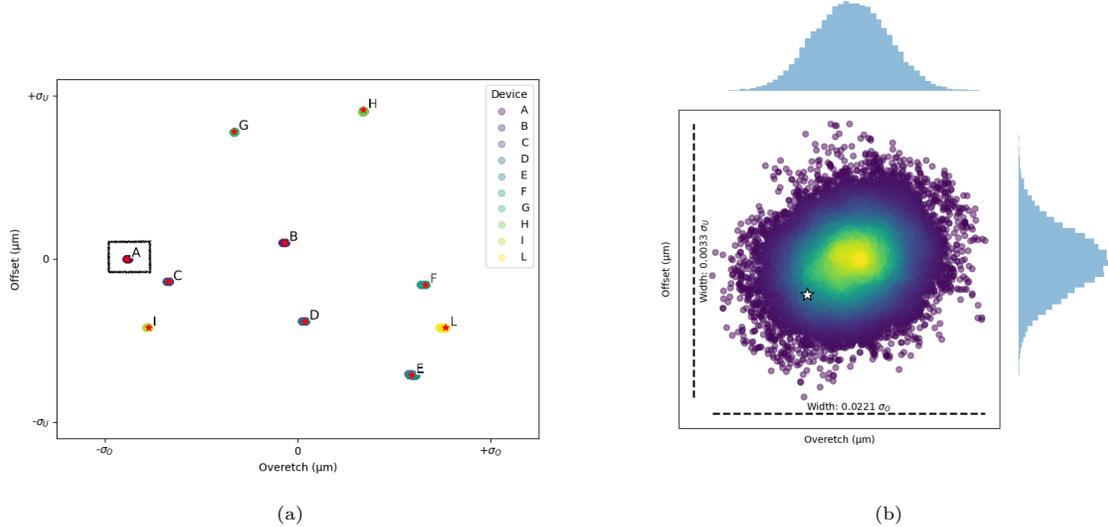


Figure 12: Posterior distribution details after launching the MCMC: (a) Final posterior distributions of the MCMC samples in the O-U domain, as obtained for ten distinct devices; (b) Scatter plot of the MCMC samples and corresponding histograms off the model parameter distributions for device A, white star represent the exact unknown parameters.

investigated the impact of various noise levels on the results. Smaller noise levels correspond to reduced errors and narrower confidence intervals, whereas higher noise levels exhibit the opposite effect. It is noteworthy that with very small noise levels, the posterior distribution may not encompass the true parameters due to the noise approaching the surrogate model’s error, resulting in minimal discrepancy. Additionally, the results demonstrate robust estimations for noise factors up to 1000.

Figure 12a presents the estimated geometric parameters, specifically overetch and offset, for the ten devices. In this figure, red stars denote the actual parameter combinations, whereas the point clouds represent the estimated posterior distributions for each case. We observe that for all devices under investigation, the estimated parameters closely align with the corresponding MCMC outputs, exhibiting in addition low variance in the distributions. Figure 12b provides an closer view of the joint distribution between overetch and offset parameters for device A, illustrating that the multi-component distribution is also characterized by a Gaussian shape, which is consistent with the assumed noise characteristics. The samples obtained for the other nine devices can be found in the Appendix. We limit ourselves to show only the overetch and offset parameters as they are the ones having a higher impact on the sensitivity.

Finally, the predictions obtained for the sensitivity of the ten devices are shown in Figure 13a, where each box-plot is representative of the MCMC output for each device and is compared to the corresponding real sensitivity of the device. We can clearly see that in every case the distribution is located very close to the real sensitivity. Using the mean of the distributions as the predictor for the sensitivity of the device yields very high accuracy, with a maximum error of 1.1% for case H, in line with industrial required accuracy. In every case, the confidence interval is very narrow, except for device H, having a total width of 4%.

We can conclude that the employed strategy provides very good performance in terms of accuracy, both in predicting the geometric parameters of the device and the quantity of interest, and proves to be a potential strategy to be used in industry for similar scenarios. We remark that finding the optimal hyper-parameters to optimize the performance of the MCMC is not trivial, with small changes highly affecting the results and case-to-case differences. As a possible alternative, we propose to rely on Hamiltonian MCMC approaches, like NUTS, to avoid this problem and accelerating the procedure, too. An additional advantage of incorporating a neural-network surrogate model in MCMC sampling [74] is its inherent capability for derivative calculation. This is achieved through backpropagation, enabling straightforward computation of gradients of the model

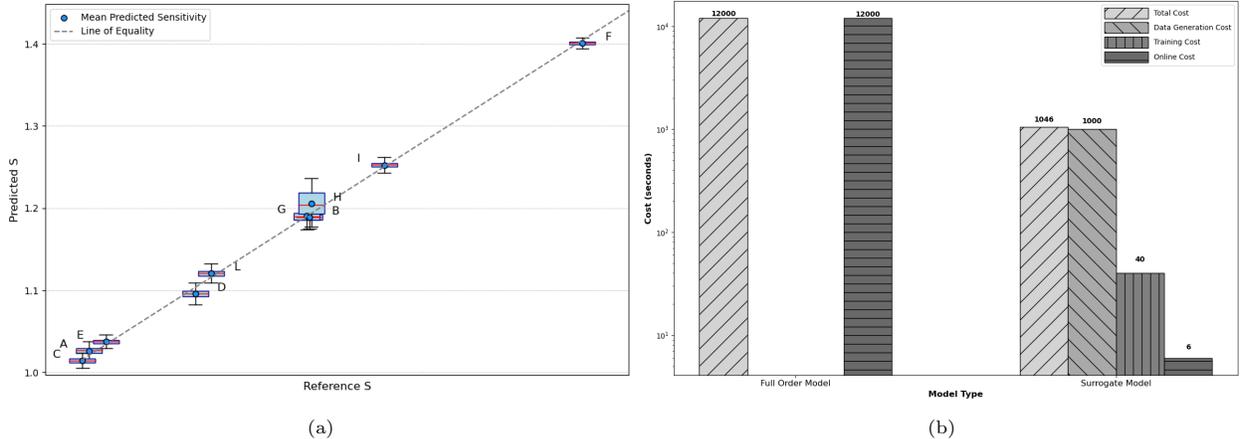


Figure 13: Performance details for the device sensitivity estimation: (a) Posterior distribution of the sensitivities for ten of the tested devices; (b) Comparison of the offline and online costs if using the high fidelity FEM model vs the data driven surrogate model. The offline costs is composed of data generation and training and is related only to the surrogate model, while the online cost is related to the MCMC procedure and the likelihood evaluation at each Metropolis-Hastings step.

with respect to input parameters. Such gradients are needed for Hamiltonian approaches. In contrast, using Finite Element Method (FEM) simulations typically necessitates an adjoint solver for derivative calculations, which is generally non-trivial and adds computational complexity. For the cases we considered, the MCMC procedure takes about 6 seconds, with a great speed-up in the online phase thanks to the use of both the surrogate models; these latter completely justify the cost of the offline data generation and training of the surrogate models. The comparison of the offline and online costs when using an high-fidelity FEM model vs when using the surrogate model based on a data-driven supervised learning strategy is presented in Fig. 13b. It is evident that even though the cost of a single high-fidelity simulation is relatively low—approximately 2 seconds for the first configuration and 1 second for the second one—the speedup provided by the surrogate model is essential during the identification phase. Since the cost of each neural network evaluation is on the order of milliseconds, the majority of the total cost of identification is attributed to the offline phase, which consists of data generation and training time. We emphasize that once this phase is completed, identification can be achieved in just a few seconds for each new instance of a device to be calibrated. Yet, different sampling strategies with respect to Metropolis Hastings should be investigated in order to reach faster convergence of the chain and obtain a further decrease the cost – such as, e.g., variational inference or Laplace approximations, just to mention two instances.

5. Conclusion

In this paper we have carried out an extensive computational characterization and uncertainty analysis of MEMS capacitive accelerometers. We have merged supervised learning-based surrogate models, specifically artificial neural networks, with conventional MCMC techniques to construct a thorough and robust characterization framework. Two distinct surrogate models have been developed within this strategy: the first model is designed to translate fabrication uncertainties into variations in capacitance over time, triggered by an active voltage signal applied across the MEMS electrodes. This model leverages data from sensors at the end of the manufacturing process, offering a cost-effective and efficient alternative to traditional device stimulation methods involving external accelerations. The second model is instead meant to correlate the manufacturing parameters directly with the device’s electro-mechanical sensitivity.

Ideally, upon the production of a new accelerometer, a new signal is captured, starting a Bayesian model update process with the initial surrogate model to deduce the posterior distribution of the device’s unknown fabrication parameters. Following the acquisition of these posterior distributions, each data point is fed into the secondary surrogate model to derive the predictive posterior distribution concerning the device’s

sensitivity. This process facilitates the extraction of sensitivity estimators from the distribution, which are used for calibrating the accelerometer’s readings. Moreover, this methodology provides a quantification of uncertainty for each parameter, thereby enhancing the interpretability of test outcomes.

The surrogate models developed in this study are trained using data obtained through high-fidelity Finite Element simulations, thus ensuring accurate predictions of MEMS capacitive accelerometer performance across various fabrication conditions. Leveraging a data-driven methodology allows us to reduce computational requirements significantly, without compromising the precision of the results, and thus enabling detailed and efficient stochastic evaluations that are orders of magnitude faster than the ones obtained with a FOM. These models have successfully estimated accelerometer sensitivities and yielded the characterization of mechanical offsets, over-etching, and thickness discrepancies. Such detailed analyses may be instrumental in estimating additional parameters of interest, enhancing the overall utility of the models in practical applications.

Future studies could expand the range of parameters examined and explore a wider variety of MEMS device configurations. An essential parameter to be included in future analyses is the device’s quality factor, which significantly affects its behavior and may exhibit considerable variability due to different fabrication pressures.

The main challenges arising in the presented investigation are the need for a large number of samples to achieve MCMC convergence and the complexities involved in identifying optimal hyper-parameters for effective chain mixing. Exploring Hamiltonian Monte Carlo techniques – as well as other alternative estimation procedures – could offer substantial improvements in this regard. Moreover, enhancing the surrogate models with more complex multi-physics components and incorporating multi-fidelity sampling strategies in the MCMC framework represent promising avenues for further advancements.

Acknowledgment: A.M. acknowledges the project “AI4MEMS” funded by Joint Research Program on “Sensor sysTEms with Advanced Materials” (STEAM)—Politecnico di Milano and STMicroelectronics S.r.l., the PRIN 2022 Project “Numerical approximation of uncertainty quantification problems for PDEs by multi-fidelity methods (UQ-FLY)” (No. 202222PACR), funded by the European Union - NextGenerationEU, and the Project “Reduced Order Modeling and Deep Learning for the real-time approximation of PDEs (DREAM)” (Starting Grant No. FIS00003154), funded by the Italian Science Fund (FIS) - Ministero dell’Università e della Ricerca.

Funding: F.Z. is supported by the Joint Research Program on “Sensor sysTEms with Advanced Materials” (STEAM)—Politecnico di Milano and STMicroelectronics S.r.l. through the project “AI4MEMS” (Artificial Intelligence for MEMS) in the topic “Virtual Twins MEMS”.

Data Availability Statement: The simulation data used during the current study is available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- [1] Alberto Corigliano, Raffaele Ardito, Claudia Comi, Attilio Frangi, Aldo Ghisi, and Stefano Mariani. *Mechanics of microsystems*. John Wiley & Sons, 2018.
- [2] Jack W Judy. Microelectromechanical systems (MEMS): fabrication, design and applications. *Smart Materials and Structures*, 10(6):1115, 2001. doi: 10.1088/0964-1726/10/6/301.
- [3] Ramin Mirzazadeh, Saeed Eftekhari Azam, and Stefano Mariani. Mechanical Characterization of Polysilicon MEMS: A Hybrid TCMC/POD-Kriging Approach. *Sensors*, 18(4):1243, 2018. doi: 10.3390/s18041243.

- [4] Yoon Shik Hong, Jong Hyun Lee, and Soo Hyun Kim. A laterally driven symmetric micro-resonator for gyroscopic applications. *Journal of Micromechanics and Microengineering*, 10(3):452–458, 2000. doi: 10.1088/0960-1317/10/3/322.
- [5] Gökçen Aslan Aydemir and Afşar Saranlı. Characterization and calibration of MEMS inertial sensors for state and parameter estimation applications. *Measurement*, 45(5):1210–1225, 2012. doi: 10.1016/j.measurement.2012.01.015.
- [6] Manuel Glueck, Dayo Oshinubi, Patrick Schopp, and Yiannos Manoli. Real-Time Autocalibration of MEMS Accelerometers. *IEEE Transactions on Instrumentation and Measurement*, 63(1):96–105, 2014. doi: 10.1109/TIM.2013.2275240.
- [7] Oliver Dürr, Po-Yu Fan, and Zong-Xian Yin. Bayesian calibration of MEMS accelerometers. *IEEE Sensors Journal*, 23(12):13319–13326, 2023. doi: 10.1109/JSEN.2023.3272907.
- [8] Aparna Harindranath and Manish Arora. A systematic review of user - conducted calibration methods for MEMS-based IMUs. *Measurement*, 225:114001, 2024. doi: 10.1016/j.measurement.2023.114001.
- [9] Attilio Frangi, Carlo Cercignani, Subrata Mukherjee, and Narayan Aluru. *Advances in Multiphysics Simulation and Experimental Testing of MEMS*. Imperial College Press, 2008. doi: 10.1142/p522.
- [10] Nitin Agarwal and N. R. Aluru. A data-driven stochastic collocation approach for uncertainty quantification in MEMS. *International Journal for Numerical Methods in Engineering*, 83(5):575–597, 2010. doi: 10.1002/nme.2844.
- [11] José Pablo Quesada-Molina and Stefano Mariani. Uncertainty quantification at the microscale: A data-driven multi-scale approach. *Engineering Proceedings*, 27(1), 2022. doi: 10.3390/ecs-a-9-13351.
- [12] Benjamin Peherstorfer, Boris Kramer, and Karen Willcox. Combining multiple surrogate models to accelerate failure probability estimation with expensive high-fidelity models. *Journal of Computational Physics*, 341:61–75, 2017. doi: <https://doi.org/10.1016/j.jcp.2017.04.012>.
- [13] Nicole Aretz-Nellesen, Martin A. Grepl, and Karen Veroy. 3D-VAR for parameterized partial differential equations: A certified reduced basis approach. *Advances in Computational Mathematics*, 45(5):2369–2400, 2019. doi: 10.1007/s10444-019-09713-w.
- [14] D. Allaire and K. Willcox. Surrogate Modeling for Uncertainty Assessment with Application to Aviation Environmental System Models. *AIAA Journal*, 48(8):1791–1803, 2010. doi: 10.2514/1.J050247.
- [15] Mengwu Guo and Jan S. Hesthaven. Reduced order modeling for nonlinear structural analysis using Gaussian process regression. *Computer Methods in Applied Mechanics and Engineering*, 341:807–826, 2018. doi: 10.1016/j.cma.2018.07.017.
- [16] Ludovica Cicci, Stefania Fresca, Mengwu Guo, Andrea Manzoni, and Paolo Zunino. Uncertainty quantification for nonlinear solid mechanics using reduced order models with Gaussian process regression. *Computers & Mathematics with Applications*, 149:1–23, 2023. doi: 10.1016/j.camwa.2023.08.016.
- [17] Robert B Gramacy. *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*. CRC press, 2020. doi: 10.1201/9780367815493.
- [18] Baskar Ganapathysubramanian and Nicholas Zabaras. Modeling diffusion in random heterogeneous media: Data-driven models, stochastic collocation and the variational multiscale method. *Journal of Computational Physics*, 226(1):326–353, 2007. doi: 10.1016/j.jcp.2007.04.009.
- [19] Patricia Astrid, Siep Weiland, Karen Willcox, and Ton Backx. Missing Point Estimation in Models Described by Proper Orthogonal Decomposition. *IEEE Transactions on Automatic Control*, 53(10): 2237–2251, 2008. doi: 10.1109/TAC.2008.2006102.

- [20] G. Kerschen, J.C. Golinval, A.F. Vakakis, and L.A. Bergman. The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: An overview. *Nonlinear Dynamics*, 41(1):147–169, 2005. doi: 10.1007/s11071-005-2803-2.
- [21] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced Basis Methods for Partial Differential Equations: An Introduction*. Springer International Publishing, 2015. doi: 10.1007/978-3-319-15431-2.
- [22] Attilio Frangi and Giorgio Gobat. Reduced order modelling of the non-linear stiffness in MEMS resonators. *International Journal of Non-Linear Mechanics*, 116:211–218, 2019. doi: 10.1016/j.ijnonlinmec.2019.07.002.
- [23] Cyril Touzé. Normal form theory and nonlinear normal modes: theoretical settings and applications. In *Modal analysis of nonlinear mechanical systems*, pages 75–160. Springer, 2014. doi: 10.1007/978-3-7091-1791-0_3.
- [24] Alberto Corigliano, Martino Dossi, and Stefano Mariani. Domain decomposition and model order reduction methods applied to the simulation of multi-physics problems in mems. *Computers & Structures*, 122:113–127, 2013. doi: <https://doi.org/10.1016/j.compstruc.2012.12.012>.
- [25] Qian Wang, Jan S. Hesthaven, and Deep Ray. Non-intrusive reduced order modeling of unsteady flows using artificial neural networks with application to a combustion problem. *Journal of Computational Physics*, 384:289–307, 2019. ISSN 0021-9991. doi: <https://doi.org/10.1016/j.jcp.2019.01.031>.
- [26] Cyril Touzé, Alessandra Vizzaccaro, and Olivier Thomas. Model order reduction methods for geometrically nonlinear structures: A review of nonlinear techniques. *Nonlinear Dynamics*, 105(2):1141–1190, 2021. doi: 10.1007/s11071-021-06693-9.
- [27] Boris Kramer and Karen E. Willcox. Nonlinear model order reduction via lifting transformations and proper orthogonal decomposition. *AIAA Journal*, 57(6):2297–2307, 2019. doi: 10.2514/1.J057791.
- [28] Annika Radermacher and Stefanie Reese. Pod-based model reduction with empirical interpolation applied to nonlinear elasticity. *International Journal for Numerical Methods in Engineering*, 107(6):477–495, 2016. doi: <https://doi.org/10.1002/nme.5177>.
- [29] Diana Bonomi, Andrea Manzoni, and Alfio Quarteroni. A matrix deim technique for model reduction of nonlinear parametrized problems in cardiac mechanics. *Computer Methods in Applied Mechanics and Engineering*, 324:300–326, 2017. doi: <https://doi.org/10.1016/j.cma.2017.06.011>.
- [30] Giorgio Gobat, Andrea Opreni, Stefania Fresca, Andrea Manzoni, and Attilio Frangi. Reduced order modeling of nonlinear microstructures through Proper Orthogonal Decomposition. *Mechanical Systems and Signal Processing*, 171:108864, 2022. doi: 10.1016/j.ymssp.2022.108864.
- [31] G. Kerschen, M. Peeters, J.C. Golinval, and A.F. Vakakis. Nonlinear normal modes, Part I: A useful framework for the structural dynamicist. *Mechanical Systems and Signal Processing*, 23(1):170–194, 2009. doi: 10.1016/j.ymssp.2008.04.002.
- [32] L. Renson, G. Kerschen, and B. Cochelin. Numerical computation of nonlinear normal modes in mechanical engineering. *Journal of Sound and Vibration*, 364:177–206, 2016. doi: 10.1016/j.jsv.2015.09.033.
- [33] Alessandra Vizzaccaro, Yichang Shen, Loïc Salles, Jiří Blahoš, and Cyril Touzé. Direct computation of nonlinear mapping via normal form for reduced-order models of finite element nonlinear structures. *Computer Methods in Applied Mechanics and Engineering*, 384:113957, 2021. doi: <https://doi.org/10.1016/j.cma.2021.113957>.
- [34] Xavier Cabré, Ernest Fontich, and Rafael de la Llave. The parameterization method for invariant manifolds III: Overview and applications. *Journal of Differential Equations*, 218(2):444–515, 2005. doi: 10.1016/j.jde.2004.12.003.

- [35] Attilio Frangi, Alessio Colombo, Alessandra Vizzaccaro, and Cyril Touzé. Reduced order modelling of fully coupled electro-mechanical systems through invariant manifolds with applications to microstructures. (arXiv:2312.14803), 2023. doi: 10.48550/arXiv.2312.14803.
- [36] Stefania Fresca and Andrea Manzoni. Pod-dl-rom: Enhancing deep learning-based reduced order models for nonlinear parametrized pdes by proper orthogonal decomposition. *Computer Methods in Applied Mechanics and Engineering*, 388:114181, 2022. doi: 10.1016/j.cma.2021.114181.
- [37] 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. doi: 10.1098/rsos.201121.
- [38] Ehsan Haghghat, Maziar Raissi, Adrian Moure, Hector Gomez, and Ruben Juanes. A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics. *Computer Methods in Applied Mechanics and Engineering*, 379:113741, 2021. doi: 10.1016/j.cma.2021.113741.
- [39] Matteo Torzoni, Andrea Manzoni, and Stefano Mariani. A multi-fidelity surrogate model for structural health monitoring exploiting model order reduction and artificial neural networks. *Mechanical Systems and Signal Processing*, 197:110376, 2023. doi: 10.1016/j.ymssp.2023.110376.
- [40] Michael McGurk, Adolphus Lye, Ludovic Renson, and Jie Yuan. Data-driven Bayesian inference for stochastic model identification of nonlinear aeroelastic systems. 2023. doi: 10.21203/rs.3.rs-3408685/v1. PREPRINT (Version 1) available at Research Square.
- [41] Stefania Fresca, Giorgio Gobat, Patrick Fedeli, Attilio Frangi, and Andrea Manzoni. Deep learning-based reduced order models for the real-time simulation of the nonlinear dynamics of microstructures. *International Journal for Numerical Methods in Engineering*, 123(20):4749–4777, 2022. doi: 10.1002/nme.7054.
- [42] Giorgio Gobat, Alessia Baronchelli, Stefania Fresca, and Attilio Frangi. Modelling the Periodic Response of Micro-Electromechanical Systems through Deep Learning-Based Approaches. *Actuators*, 12(7):278, 2023. doi: 10.3390/act12070278.
- [43] Giorgio Gobat, Stefania Fresca, Andrea Manzoni, and Attilio Frangi. Reduced Order Modeling of Nonlinear Vibrating Multiphysics Microstructures with Deep Learning-Based Approaches. *Sensors*, 23(6):3001, 2023. doi: 10.3390/s23063001.
- [44] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences of the United States of America*, 113(15):3932–3937, 2016. doi: 10.1073/pnas.1517384113.
- [45] Paolo Conti, Giorgio Gobat, Stefania Fresca, Andrea Manzoni, and Attilio Frangi. Reduced order modeling of parametrized systems through autoencoders and SINDy approach: Continuation of periodic solutions. *Computer Methods in Applied Mechanics and Engineering*, 411:116072, 2023. doi: 10.1016/j.cma.2023.116072.
- [46] Y.C. Wang, S.G. Adams, J.S. Thorp, N.C. MacDonald, P. Hartwell, and F. Bertsch. Chaos in MEMS, parameter estimation and its potential application. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 45(10):1013–1020, 1998. doi: 10.1109/81.728856.
- [47] Charles J. Geyer. Introduction to Markov Chain Monte Carlo. In *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, 2011. doi: 10.1201/b10905-2.
- [48] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*, volume 160 of *Applied Mathematical Sciences*. Springer Science+Business Media, Inc., 2005. doi: 10.1007/b138659.
- [49] Albert Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, 2005. doi: 10.1137/1.9780898717921.

- [50] Ching Jianye and Chen Yi-Chu. Transitional Markov Chain Monte Carlo Method for Bayesian Model Updating, Model Class Selection, and Model Averaging. *Journal of Engineering Mechanics*, 133(7): 816–832, 2007. doi: 10.1061/(ASCE)0733-9399(2007)133:7(816).
- [51] Radford Neal. MCMC Using Hamiltonian Dynamics. In Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng, editors, *Handbook of Markov Chain Monte Carlo*, volume 20116022. Chapman and Hall/CRC, 2011. doi: 10.1201/b10905-6.
- [52] Matthew D. Hoffman and Andrew Gelman. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(47):1593–1623, 2014.
- [53] Michael B. Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24:259–328, 2015. doi: 10.1017/S096249291500001X.
- [54] C. Soize and C. Farhat. Probabilistic learning for modeling and quantifying model-form uncertainties in nonlinear computational mechanics. *International Journal for Numerical Methods in Engineering*, 117(7):819–843, 2019. doi: <https://doi.org/10.1002/nme.5980>.
- [55] Moritz Berger, Christian Schott, and Oliver Paul. Bayesian sensor calibration. *IEEE Sensors Journal*, 22(20):19384–19399, 2022. doi: 10.1109/JSEN.2022.3199485.
- [56] Shiwei Lan, Shuyi Li, and Babak Shahbaba. Scaling up bayesian uncertainty quantification for inverse problems using deep neural networks. *SIAM/ASA Journal on Uncertainty Quantification*, 10(4):1684–1713, 2022. doi: 10.1137/21M1439456.
- [57] Dhruv V. Patel, Deep Ray, and Assad A. Oberai. Solution of physics-based bayesian inverse problems with deep generative priors. *Computer Methods in Applied Mechanics and Engineering*, 400:115428, 2022. ISSN 0045-7825. doi: <https://doi.org/10.1016/j.cma.2022.115428>.
- [58] Audrey Olivier, Michael D. Shields, and Lori Graham-Brady. Bayesian neural networks for uncertainty quantification in data-driven materials modeling. *Computer Methods in Applied Mechanics and Engineering*, 386:114079, 2021. doi: <https://doi.org/10.1016/j.cma.2021.114079>.
- [59] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8):2554–2558, 1982.
- [60] Stephen Timoshenko. *History of strength of materials: with a brief account of the history of theory of elasticity and theory of structures*. Courier Corporation, 1983.
- [61] Harlan B. Palmer. The capacitance of a parallel-plate capacitor by the Schwartz-Christoffel transformation. *Electrical Engineering*, 56(3):363–368, 1937. doi: 10.1109/EE.1937.6540485.
- [62] Xingguo Xiong, Yu-Liang Wu, and W.-B. Jone. A dual-mode built-in self-test technique for capacitive MEMS devices. *IEEE Transactions on Instrumentation and Measurement*, 54(5):1739–1750, 2005. doi: 10.1109/TIM.2005.855094.
- [63] J. S. Hesthaven and S. Ubbiali. Non-intrusive reduced order modeling of nonlinear problems using neural networks. *Journal of Computational Physics*, 363:55–78, 2018. doi: 10.1016/j.jcp.2018.02.037.
- [64] Kuan Lu, Yulin Jin, Yushu Chen, Yongfeng Yang, Lei Hou, Zhiyong Zhang, Zhonggang Li, and Chao Fu. Review for order reduction based on proper orthogonal decomposition and outlooks of applications in mechanical systems. *Mechanical Systems and Signal Processing*, 123:264–297, 2019. doi: <https://doi.org/10.1016/j.ymssp.2019.01.018>.
- [65] G. Kerschen and J.C. Golinval. Physical interpretation of the proper orthogonal modes using the singular value decomposition. *Journal of Sound and Vibration*, 249(5):849–865, 2002. doi: 10.1006/jsvi.2001.3930.

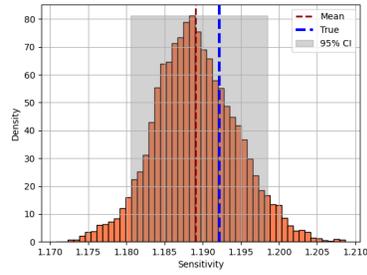
- [66] Ralph C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2013. doi: 10.1137/1.9781611973228.
- [67] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 04 1970. doi: 10.1093/biomet/57.1.97.
- [68] Lam Research. CoventorMP. <https://www.lamresearch.com/coventormp/>.
- [69] François Chollet et al. Keras. <https://keras.io>, 2015.
- [70] Mark Ebden. *Gaussian Processes: A Quick Introduction*, 2015.
- [71] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8): 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.
- [72] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved \hat{R} for assessing convergence of MCMC. *Bayesian Analysis*, 16(2), 2021. doi: 10.1214/20-BA1221.
- [73] Nicolai A. B. Riis, Amal M. A. Alghamdi, Felipe Uribe, Silja L. Christensen, Babak M. Afkham, Per Christian Hansen, and Jakob S. Jørgensen. CUQIpy – Part I: Computational uncertainty quantification for inverse problems in Python, 2023.
- [74] Teo Deveney, Eike H. Mueller, and Tony Shardlow. Deep surrogate accelerated delayed-acceptance hamiltonian monte carlo for bayesian inference of spatio-temporal heat fluxes in rotating disc systems. *SIAM/ASA Journal on Uncertainty Quantification*, 11(3):970–995, 2023. doi: 10.1137/22M1513113.

Appendix A. Comprehensive Posterior Distribution Results

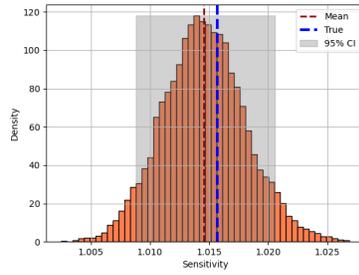
The following figures present comprehensive posterior distribution results obtained from the Markov Chain Monte Carlo (MCMC) analysis for devices B to L. These results complement the ones presented in the main body of the paper, where only results for one device were included for simplicity. The additional results demonstrate the robustness of the proposed approach under different testing conditions.

Figure A.14 shows the final results of the posterior distribution of the sensitivity obtained for 9 additional devices. The plots show consistent results with those outlined in the paper, characterized by narrow confidence intervals and small errors. One exception is device "H," which displays a more irregular shape of the distribution, though it still has small errors. This irregularity is likely due to high self-correlation of the samples, which can be mitigated by increasing sample size and thinning out samples.

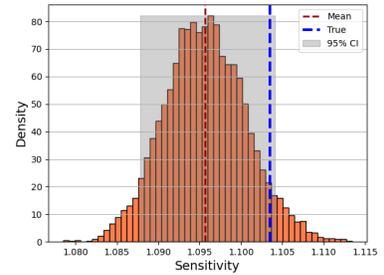
Figure A.15 shows the scatter plots of the posterior distributions of the overetch and offset parameters for devices B to L. In this case, the cloud of points is concentrated very close to the true unknown parameters and generally has a Gaussian and regular shape, often characterized by non-null correlation between the two parameters. Device "H" again shows a more irregular pattern, for which the same considerations as before apply.



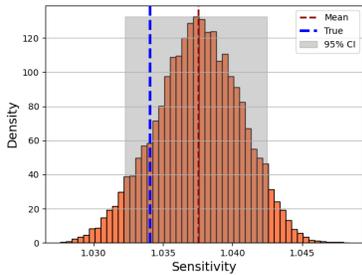
(B) Device B



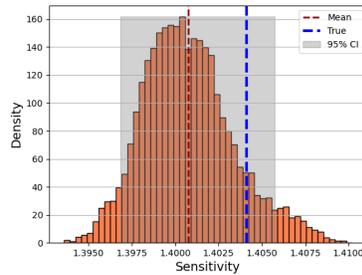
(C) Device C



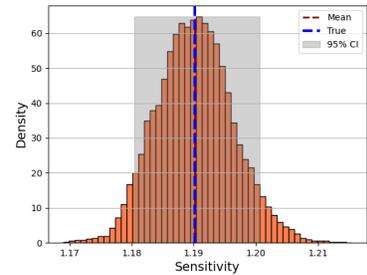
(D) Device D



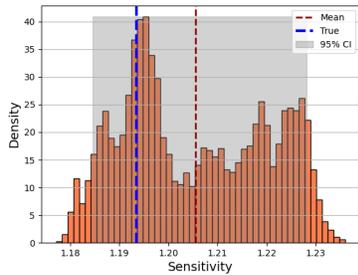
(E) Device E



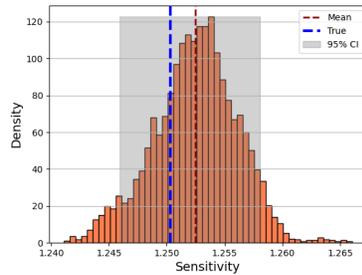
(F) Device F



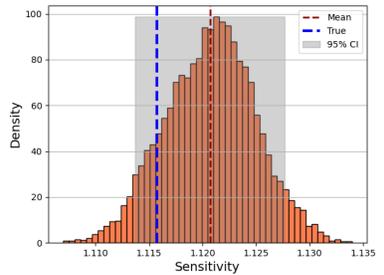
(G) Device G



(H) Device H



(I) Device I



(L) Device L

Figure A.14: Histograms of the posterior distribution of sensitivity for devices B to L. Each subplot shows the distribution of sensitivity values inferred from the MCMC analysis.

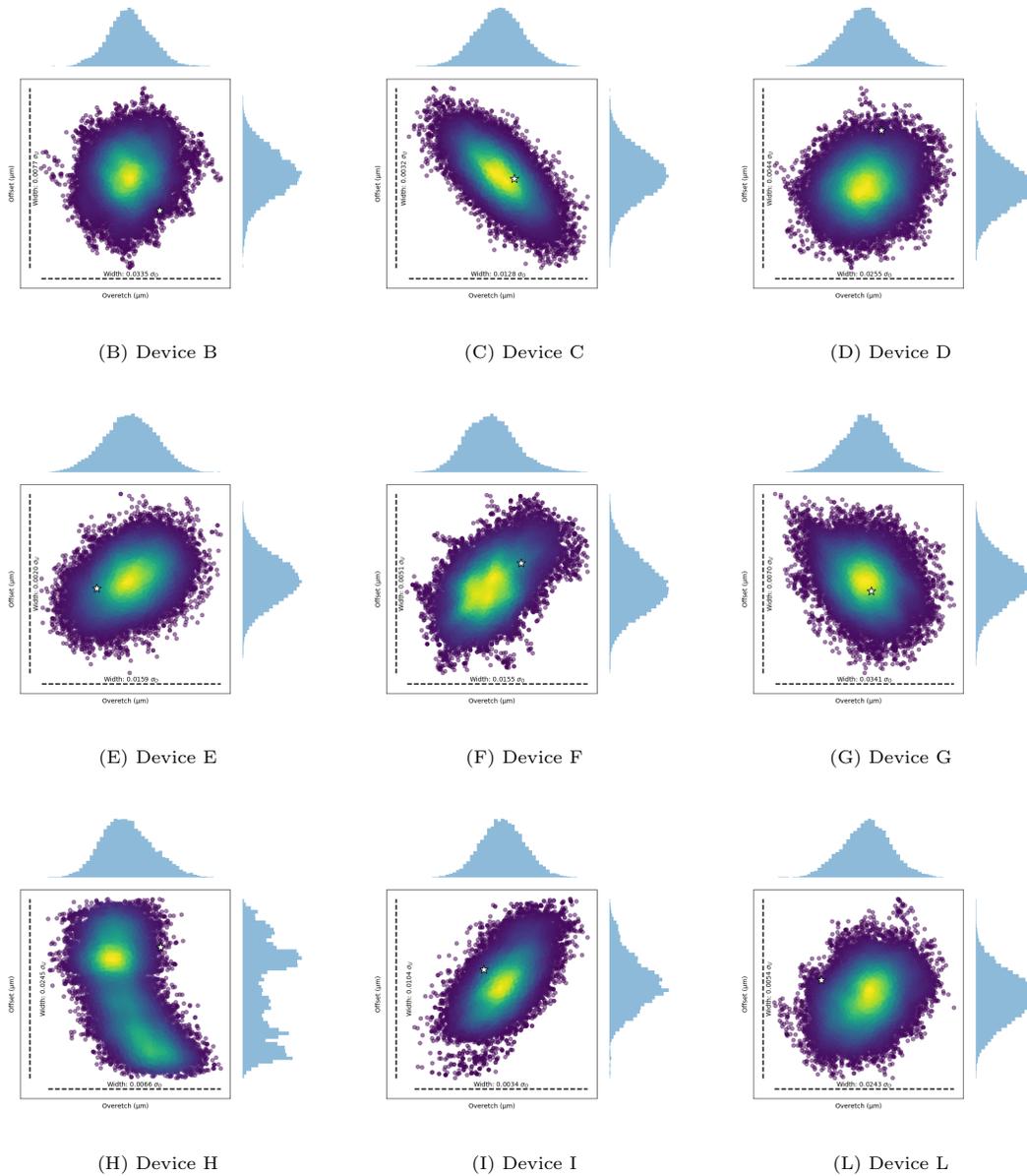


Figure A.15: Scatter plots of the multi-component posterior distribution of overetch and offset for devices B to L. Each subplot shows how overetch and offset values covary in the posterior distribution, with the white star indicating the real combination observed.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

- 94/2024** Franco, N.R.; Fresca, S.; Tombari, F.; Manzoni, A.
Deep Learning-based surrogate models for parametrized PDEs: handling geometric variability through graph neural networks
- 93/2024** Conti, P.; Kneifl, J.; Manzoni, A.; Frangi, A.; Fehr, J.; Brunton, S.L.; Kutz, J.N.
VENI, VINDy, VICI - a variational reduced-order modeling framework with uncertainty quantification
- 91/2024** Ciaramella, G.; Kartmann, M.; Mueller, G.
Solving Semi-Linear Elliptic Optimal Control Problems with L1-Cost via Regularization and RAS-Preconditioned Newton Methods
- 86/2024** Franco, N.R.; Fraulin, D.; Manzoni, A.; Zunino, P.
On the latent dimension of deep autoencoders for reduced order modeling of PDEs parametrized by random fields
- Castiglionea, C.; Arnonec, E.; Bernardi, M.; Farcomeni, A.; Sangalli, L.M.
PDE-regularised spatial quantile regression
- 88/2024** Regazzoni, F.; Poggesi, C.; Ferrantini, C.
Elucidating the cellular determinants of the end-systolic pressure-volume relationship of the heart via computational modelling
- 85/2024** Brivio, S.; Franco, Nicola R.; Fresca, S.; Manzoni, A.
Error estimates for POD-DL-ROMs: a deep learning framework for reduced order modeling of nonlinear parametrized PDEs enhanced by proper orthogonal decomposition
- 90/2024** Tomasetto, M.; Arnone, E.; Sangalli, L.M.
Modeling anisotropy and non-stationarity through physics-informed spatial regression
- 83/2024** Conti, P.; Guo, M.; Manzoni, A.; Frangi, A.; Brunton, S. L.; Kutz, J.N.
Multi-fidelity reduced-order surrogate modelling
- 82/2024** Rosafalco, L.; Conti, P.; Manzoni, A.; Mariani, S.; Frangi, A.
EKF-SINDy: Empowering the extended Kalman filter with sparse identification of nonlinear dynamics