
GAUSSIAN PROCESS REGRESSION FOR EFFICIENT DETECTION OF FABRICATION INACCURACIES IN MICROSENSORS



POLITECNICO
MILANO 1863



FILIPPO ZACCHEI

APRIL 2025

PHD COURSE: ADVANCED STATISTICAL METHODS FOR COMPLEX DATA

PROF. FRANCESCA IEVA

OUTLINE

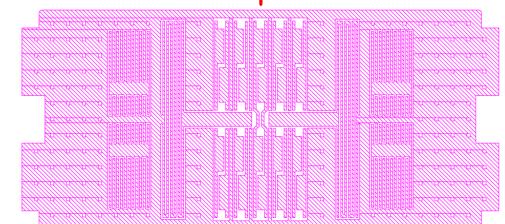
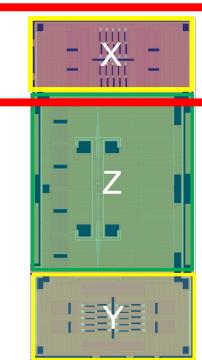
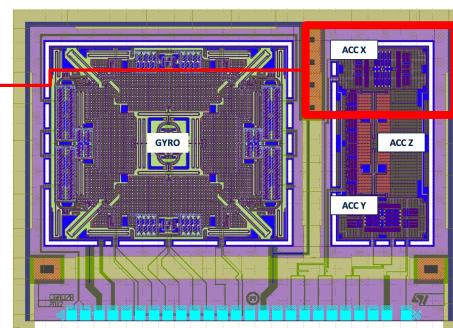
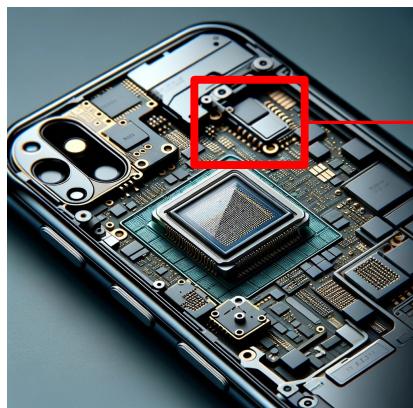
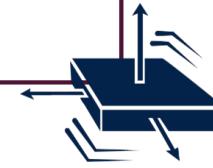
- Context and Motivation
- Gaussian Process Regression
- A Comparison with Kriging
- GPR for Online Learning
- Numerical Results

Objective

Calibration of MEMS accelerometers:

- Fabrication may lead geometrical layout to differ up to 10% w.r.t. the intended one.
- The sensitivity of the device w.r.t. external accelerations is highly impacted by fabrication inaccuracies.
- Sensitivity is generally estimated using labor intensive mechanical tests.

Calibration of MEMS accelerometer using cheap electronic tests



GENERAL FRAMEWORK - METROPOLIS HASTINGS

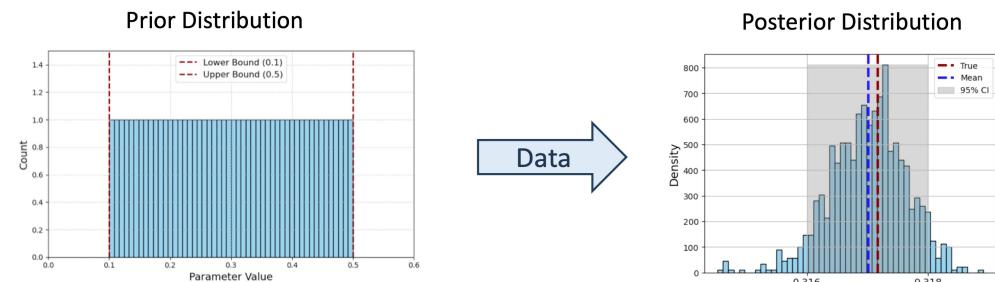
- Measurements are stored in $y^{\text{exp}} \in R^d$.
- Uncertain parameters are stored in $x \in R^p$
- The two are connected by a **forward model** f_{HF} :

$$y^{\text{exp}} = f_{HF}(x) + \varepsilon, \quad \varepsilon \sim N(0, \Sigma)$$

- Goal: infer the **posterior distribution** of x :

$$\pi(x | y^{\text{exp}}) = \frac{\pi(y^{\text{exp}} | x) \pi(x)}{\pi(y^{\text{exp}})}$$

- Standard approach: Metropolis-Hastings



GENERAL FRAMEWORK - METROPOLIS HASTINGS

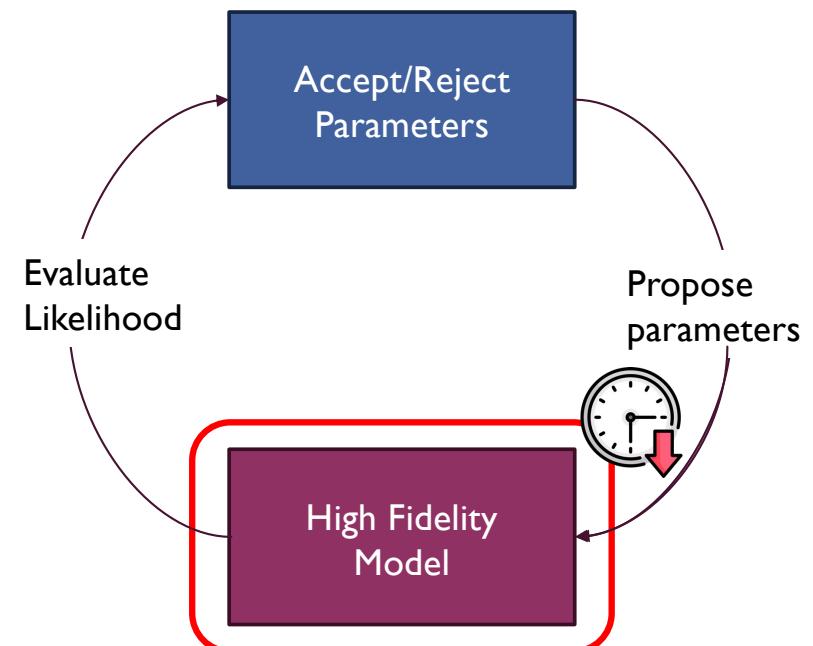
- Measurements are stored in $y^{\text{exp}} \in R^d$.
- Uncertain parameters are stored in $x \in R^p$
- The two are connected by a **forward model** f_{HF} :

$$y^{\text{exp}} = f_{HF}(x) + \varepsilon, \quad \varepsilon \sim N(0, \Sigma)$$

- Goal: infer the **posterior distribution** of x :

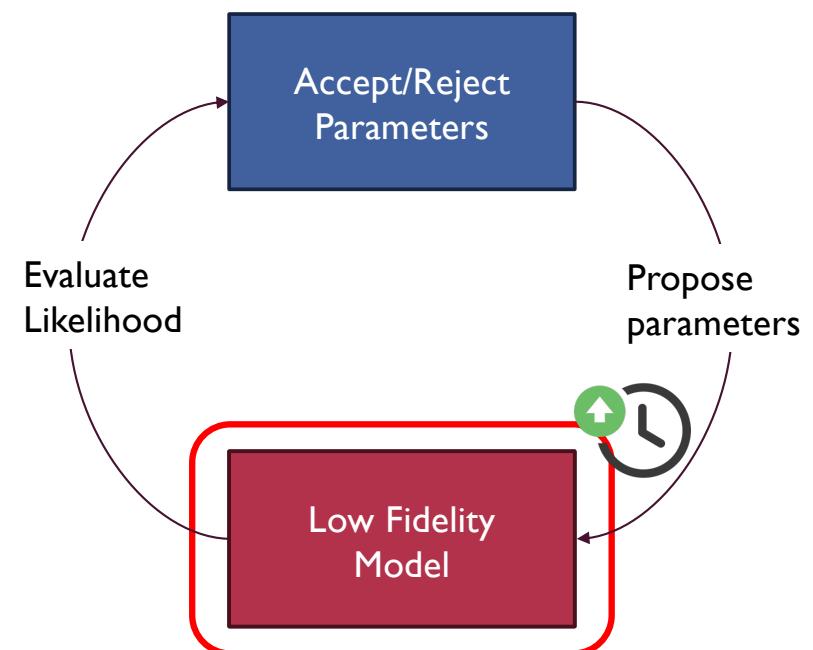
$$\pi(\mu | y^{\text{exp}}) = \frac{\pi(y^{\text{exp}} | x) \pi(x)}{\pi(y^{\text{exp}})}$$

- Standard approach: Metropolis-Hastings



GENERAL FRAMEWORK - LOW FIDELITY

- Replace the forward model with a low fidelity approximation f_{LF} :
$$y^{\text{exp}} \approx f_{LF}(x) + \varepsilon, \quad \varepsilon \sim N(0, \Sigma)$$
- Huge gains in sampling efficiency
- The surrogate determines the accuracy of the estimate



Zacchei, Filippo, et al. "Neural networks based surrogate modeling for efficient uncertainty quantification and calibration of MEMS accelerometers." *International Journal of Non-Linear Mechanics* 167 (2024): 104902.

SURROGATE MODELING

- Surrogate models replace computationally expensive forward models, enabling efficient Bayesian inference.
- This transforms the inverse problem into a **statistical learning task**, where the surrogate approximates the high-fidelity model behavior.

Gaussian Process Regression!

Key Considerations:

Error Quantification:

- We must estimate surrogate errors **accurately**, as they propagate into the **posterior distribution** and may bias inference.

Data Efficiency:

- Surrogates must be trained on **fewer samples** than needed for direct MCMC, making **data efficiency** essential.

GAUSSIAN PROCESS REGRESSION

A **Gaussian Process** is a distribution over functions:

$$f(x) \sim GP(m(x), k(x, x'))$$

- $m(x)$: mean function (often assumed zero)
- $k(x, x')$: covariance (kernel) function

🔍 Gaussian Process Regression

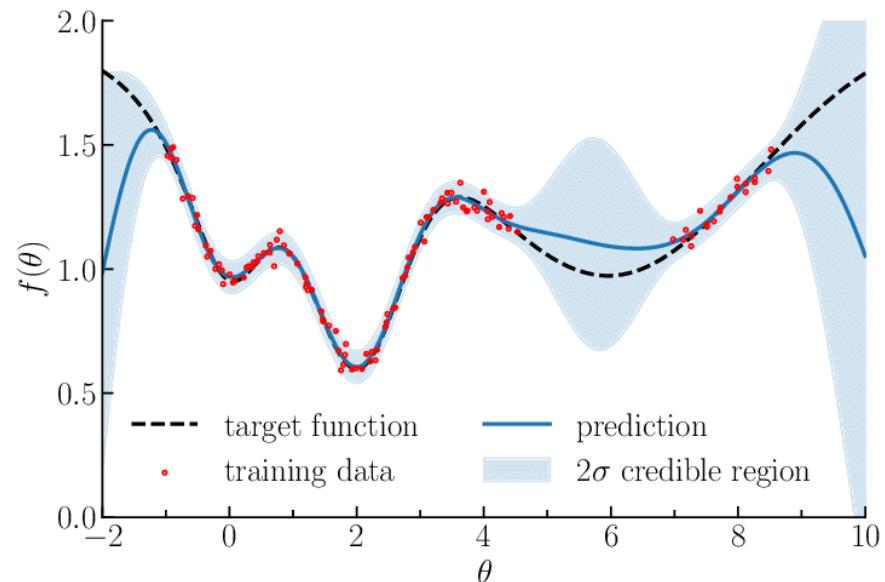
Given training data $D = \{X, y\}$ and observation noise σ_n^2 ,
the GP posterior at a test point x^* is:

Mean prediction:

$$\mu(x^*) = k^{*T} (K + \sigma_n^2 I)^{-1} y$$

Predictive variance:

$$\sigma^2(x^*) = k(x^*, x^*) - k^{*T} (K + \sigma_n^2 I)^{-1} k^*$$



Simple Kriging (SK): Mathematical Formulation

Consider a random field $Z(\mathbf{x})$, with known mean $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$. We have observations:

$$Y(\mathbf{x}_i) = Z(\mathbf{x}_i) + \varepsilon_i, \quad i = 1, \dots, n$$

The **Best Linear Unbiased Predictor (BLUP)** estimator is given by:

$$\hat{Z}(\mathbf{x}_*) = \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i) + \lambda_0$$

Minimizing the variance of prediction error subject to unbiasedness constraint yields the SK solution:

$$\hat{\boldsymbol{\lambda}} = \Sigma^{-1} \mathbf{k}_*, \quad \hat{\lambda}_0 = m(\mathbf{x}_*) - \hat{\boldsymbol{\lambda}}^\top \mathbf{m}$$

where: - $\Sigma = [k(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1}^n$ - $\mathbf{k}_* = [k(\mathbf{x}_1, \mathbf{x}_*), \dots, k(\mathbf{x}_n, \mathbf{x}_*)]^\top$

The prediction becomes explicitly:

$$\hat{Z}(\mathbf{x}_*) = m(\mathbf{x}_*) + \mathbf{k}_*^\top \Sigma^{-1} (Y - \mathbf{m})$$

with variance of prediction error:

$$\mathbb{V}[\hat{Z}(\mathbf{x}_*) - Z(\mathbf{x}_*)] = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^\top \Sigma^{-1} \mathbf{k}_*$$

Connection: Simple Kriging and Gaussian Process Regression

Gaussian Process Regression (GPR) explicitly assumes a joint Gaussian distribution for observed values and predictions:

$$\begin{bmatrix} \mathbf{Z} \\ Z(\mathbf{x}_*) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(\mathbf{X}) \\ m(\mathbf{x}_*) \end{bmatrix}, \begin{bmatrix} \Sigma & \mathbf{k}_* \\ \mathbf{k}_*^\top & k(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix} \right)$$

The GPR posterior predictive distribution at a new location \mathbf{x}_* is:

$$Z(\mathbf{x}_*) | \mathbf{Z}, \mathbf{X}, \mathbf{x}_* \sim \mathcal{N}(\mu_*, \sigma_*^2)$$

with posterior mean and variance:

$$\mu_* = m(\mathbf{x}_*) + \mathbf{k}_*^\top \Sigma^{-1} (\mathbf{Z} - \mathbf{m})$$

$$\sigma_*^2 = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^\top \Sigma^{-1} \mathbf{k}_*$$

Key Observations:

- The SK estimator coincides exactly with the posterior mean of GPR.
- The SK prediction error variance matches exactly the posterior variance in GPR. - Thus, GPR generalizes SK by providing not only a point estimator (BLUP), but a full posterior predictive distribution.

Accuracy Results

Corollary 1 Suppose $U \subseteq \mathbb{R}^{d_u}$, $\Phi \in H_k(U)$, and the random surrogate model is constructed by applying Gaussian process regression to Φ , resulting in $\Phi_N \sim GP(m_N^\Phi(u), k_N(u, u'))$. Then there exist constants $C_{Co}, C'_{Co} > 0$, independent of N , such that

$$d_{\text{Hell}}(\mu^y, \mu_{\text{mean}}^{y,N}) \leq C_{Co} \|\Phi - m_N^\Phi\|_{L_{\mu^y}^2(U)},$$

Remark: accuracy is optimal when our sampling points are close to the posterior!

Theorem 3 Suppose U is a bounded Lipschitz domain, $H_k(U)$ is isomorphic to the Sobolev space $H^\tau(U)$, and $f \in H^\tau(U)$. Further suppose that for all $N \in \mathbb{N}$,

- (i) $U_N \subseteq \mathbb{R}^{d_u}$ is compact and $U_N \subseteq \{u \in \mathbb{R}^{d_u} : \pi^y(u) \leq C_1^2 N^{-\frac{2\tau}{d_u}}\}$,
- (ii) the training points D_N are sampled i.i.d. from a measure ν_N with density ρ_N satisfying $\rho_N(u) \geq \rho_{\min} > 0$ for all $u \in U \setminus U_N$, and $\rho_N(u) = 0$ otherwise,
- (iii) $U \setminus U_N$ is a Lipschitz domain that satisfies an interior cone condition with angle θ , and $U \setminus U_N$ is contained in the cube $B(u_c, R_c^N) = \{u \in \mathbb{R}^{d_u} : \|u - u_c\|_\infty \leq R_c^N\}$, for some $u_c \in \mathbb{R}^{d_u}$ and $0 < R_c^N < C_2 \log N$.

Then there exists a constant $C_{\text{Thm}} > 0$, independent of f and N , such that for all $0 \leq \beta \leq \tau$ and $\varepsilon > 0$ we have

$$\mathbb{E}_{\nu_N} \left[\|f - m_N^f\|_{H_{\mu^y}^\beta(U)} \right] \leq C_{\text{Thm}} N^{-\frac{\tau-\beta}{d_u} + \varepsilon} \|f\|_{H^\tau(U)}.$$

Furthermore, for any partitioning $U \setminus U_N \subseteq \cup_{i=1}^r B_i$, where each B_i is a bounded Lipschitz domain that satisfies an interior cone condition with angle θ' , there exists a constant C'_{Thm} such that for all $0 \leq \beta \leq \tau$ we have

$$\begin{aligned} & \mathbb{E}_{\nu_N} \left[\|f - m_N^f\|_{H_{\mu^y}^\beta(U)} \mathbf{I}_{\{h_{D_N, B_i} \leq h_0(B_i), 1 \leq i \leq n\}} \right] \\ & \leq C'_{\text{Thm}} \left(\left(\sup_{u \in U_N} \pi^y(u) \right)^{\frac{1}{2}} + \sum_{i=1}^r \left(\sup_{u \in B_i} \pi^y(u) \right)^{\frac{1}{2}} \mathbb{E}_{\nu_N} [h_{D_N, B_i}^{\tau-\beta}] \right). \end{aligned}$$

Online Gaussian Process Regression

Incremental Update (Woodbury Identity):

Given previous covariance matrix inverse K_{XX}^{-1} , adding new points X' :

$$K_{X,X}^{-1} \rightarrow \begin{bmatrix} K_{XX} & K_{XX'} \\ K_{X'X} & K_{X'X'} \end{bmatrix}^{-1}$$

is updated efficiently by:

$$\begin{aligned} \text{Set } B &= K_{XX'}, \quad C = K_{X'X}, \quad D = K_{X'X'} \\ \Rightarrow & \begin{bmatrix} K_{XX}^{-1} + K_{XX}^{-1}B\Lambda CK_{XX}^{-1} & -K_{XX}^{-1}B\Lambda \\ -\Lambda CK_{XX}^{-1} & \Lambda \end{bmatrix} \end{aligned}$$

where $\Lambda = (D - CK_{XX}^{-1}B)^{-1}$.

Computational Complexity:

- ▶ Naive GP inversion: $\mathcal{O}(N^3)$
- ▶ Online (incremental) update complexity: $\mathcal{O}(N^2)$ per new point

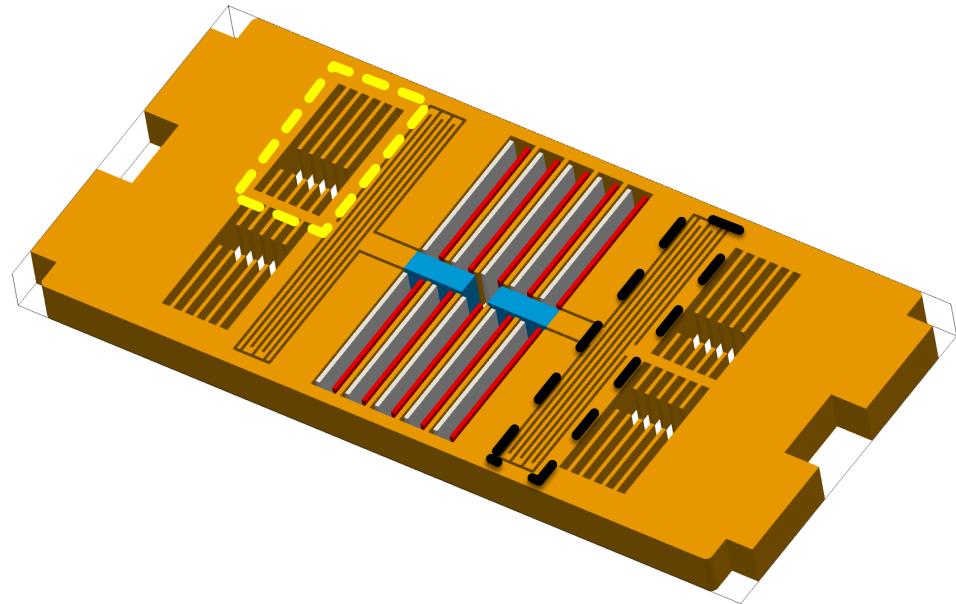
Our goal is to train the GP model during the MCMC!

NUMERICAL RESULTS



3D Model

Components	
	Shuttle mass I
	Left Electrodes
	Right Electrodes
	Anchors
	Folded Beams = Springs
	Air damper



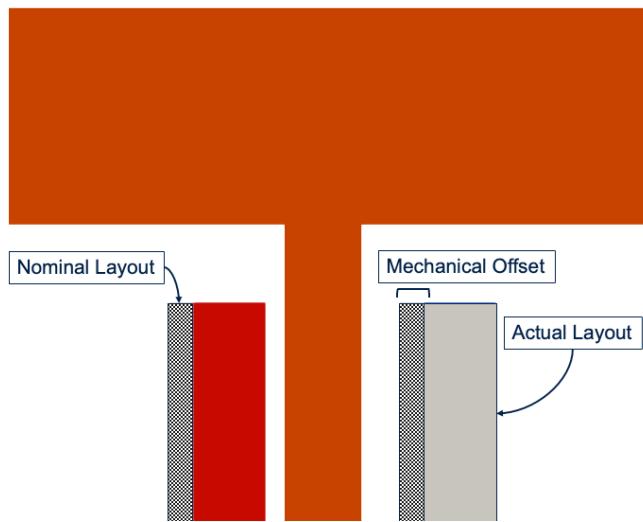
Mathematical Model

- Euler-Bernoulli elements.
- Conformal mapping for electrostatic force
- Geometric non-linearity is included.
- Small-strain assumption.
- μ embeds geometrical uncertainties

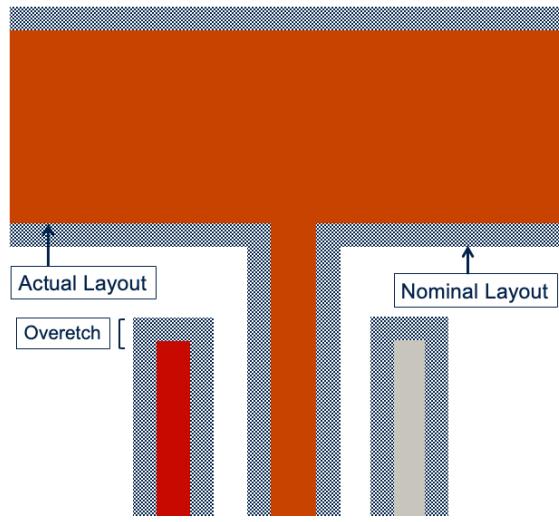
$$\begin{aligned}
 \rho_0 \ddot{\mathbf{u}}(\mathbf{X}, t; \boldsymbol{\mu}) + \mathbf{C} \dot{\mathbf{u}}(\mathbf{X}, t; \boldsymbol{\mu}) - \nabla_{\mathbf{X}} \cdot \mathbf{P}(\mathbf{u}(\mathbf{X}, t; \boldsymbol{\mu}); \boldsymbol{\mu}) &= -\rho_0 \mathbf{a}_0 && \text{in } \Omega_0 \times \mathcal{T}, \\
 \mathbf{P}(\mathbf{u}(\mathbf{X}, t; \boldsymbol{\mu}); \boldsymbol{\mu}) \cdot \mathbf{N}(\mathbf{X}) &= \mathbf{f}_{elec}(\mathbf{X}) && \text{on } \partial\Omega_{0N} \times \mathcal{T}, \\
 \mathbf{u}(\mathbf{X}, t) &= \mathbf{0} && \text{on } \partial\Omega_{0D} \times \mathcal{T}, \\
 \mathbf{u}(\mathbf{X}, 0) &= \mathbf{0} && \text{in } \Omega_0, \\
 \dot{\mathbf{u}}(\mathbf{X}, 0) &= \mathbf{0} && \text{in } \Omega_0, \\
 \operatorname{div}(\operatorname{grad}\phi(\mathbf{x})) &= 0 && \text{in } \Omega_\infty \setminus \Omega \times \mathcal{T}, \\
 \phi(\mathbf{x}) &= V_k(t) && \text{on } \partial\Omega_k, \\
 \operatorname{grad}\phi(\mathbf{x}) \cdot \mathbf{n} &= 0 && \text{on } S_\infty.
 \end{aligned}$$

Fabrication Inaccuracies

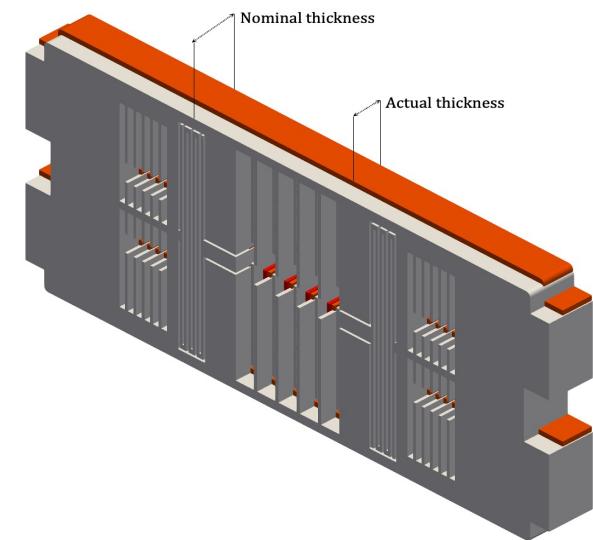
1: Mechanical Offset



2: Unilateral Overetch



3: Epipoly Thickness

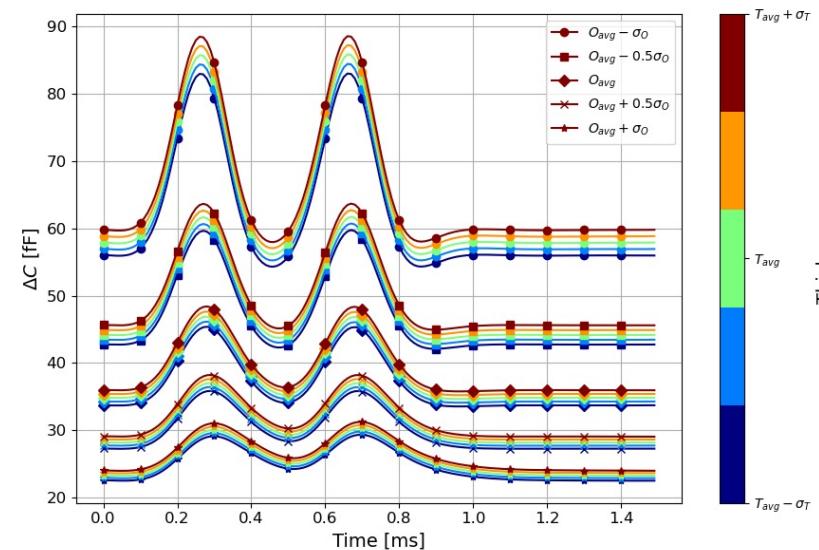
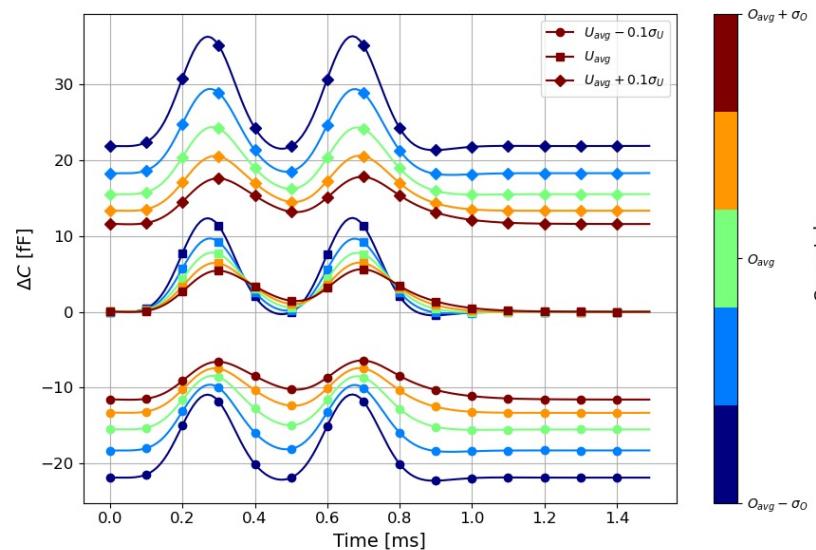
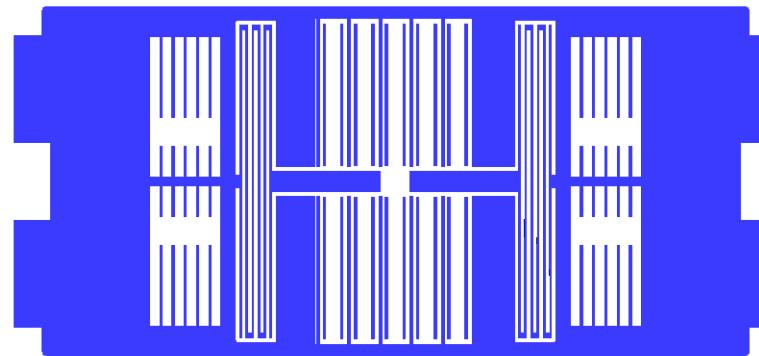


Zacchei, Filippo, et al. "Neural networks based surrogate modeling for efficient uncertainty quantification and calibration of MEMS accelerometers." *International Journal of Non-Linear Mechanics* 167 (2024): 104902.

Simulation Setup

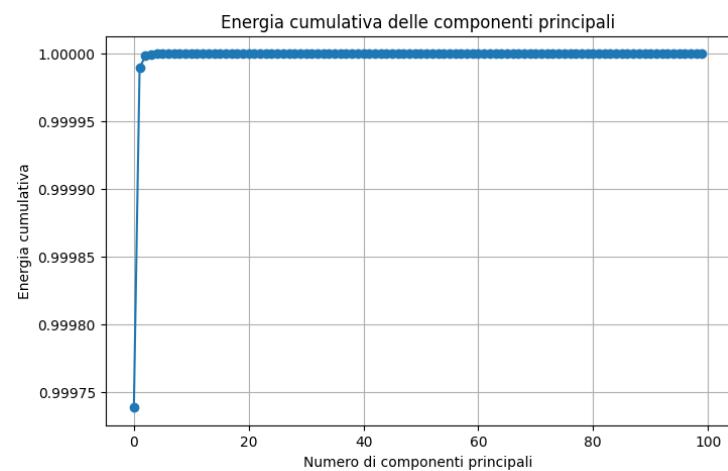
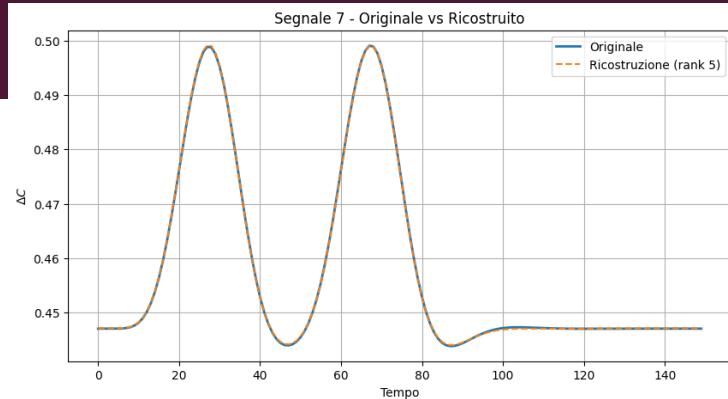
Electronic Test:

$$\begin{aligned}
 V_{right} &= 0.9 (1 + \cos(2\pi ft)) \text{ [V]} & \text{for } 0 \leq t \leq 2T \text{ [s]}, \\
 V_{left} &= 0 \text{ [V]} & \text{for } t \geq 0 \text{ [s]}, \\
 V_{right} &= 0 \text{ [V]} & \text{for } t > 2T \text{ [s]}
 \end{aligned}$$



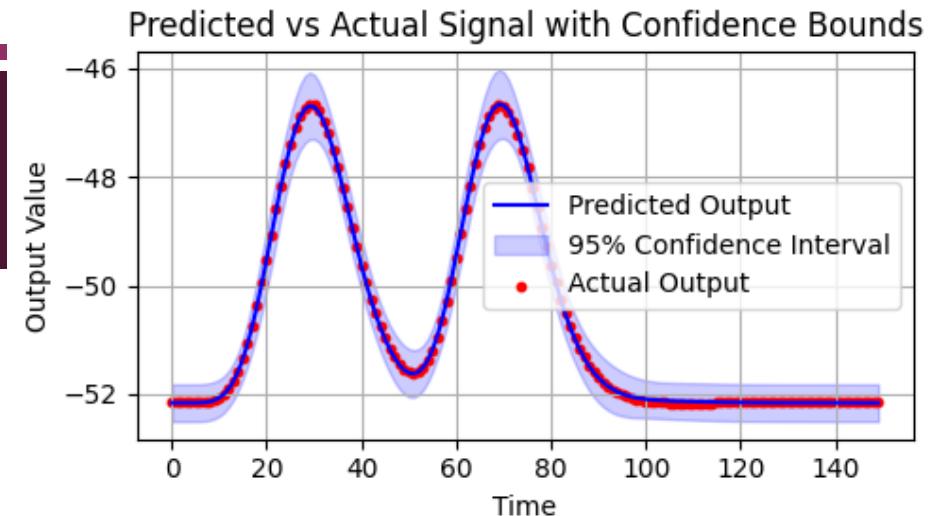
DIMENSIONALITY REDUCTION: POD-GP

- In order to have a surrogate for the time series, we decompose the signal using SVD.
- We then train an independent GP to approximate each POD mode coefficient.
- The POD guarantees that the coefficients of distinct POD modes are independent, thereby avoiding the need of a multi-output GP.



ACCURACY

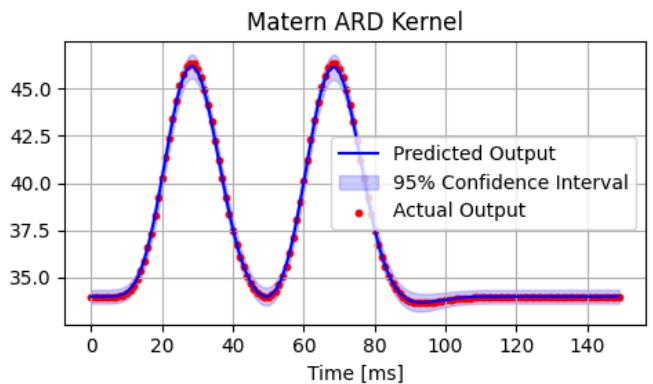
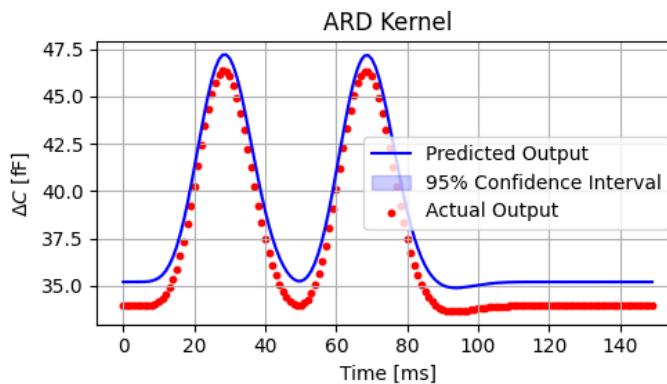
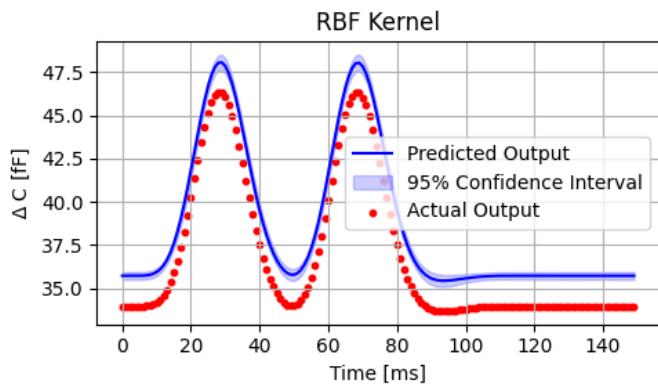
- High accuracy is obtained when increasing the number of samples.
- Our goal is to minimize the number of training data online: therefore we will stop at 100.



COVARIANCE KERNEL ROLE

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\ell^2}\right)$$

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2} \sum_{d=1}^D \frac{(x_d - x'_d)^2}{\ell_d^2}\right)$$



$$k_{5/2}(\mathbf{x}, \mathbf{x}') = \left(1 + \sqrt{5r^2(\mathbf{x}, \mathbf{x}')} + \frac{5}{3}r^2(\mathbf{x}, \mathbf{x}')\right) \exp\left(-\sqrt{5r^2(\mathbf{x}, \mathbf{x}')}\right)$$

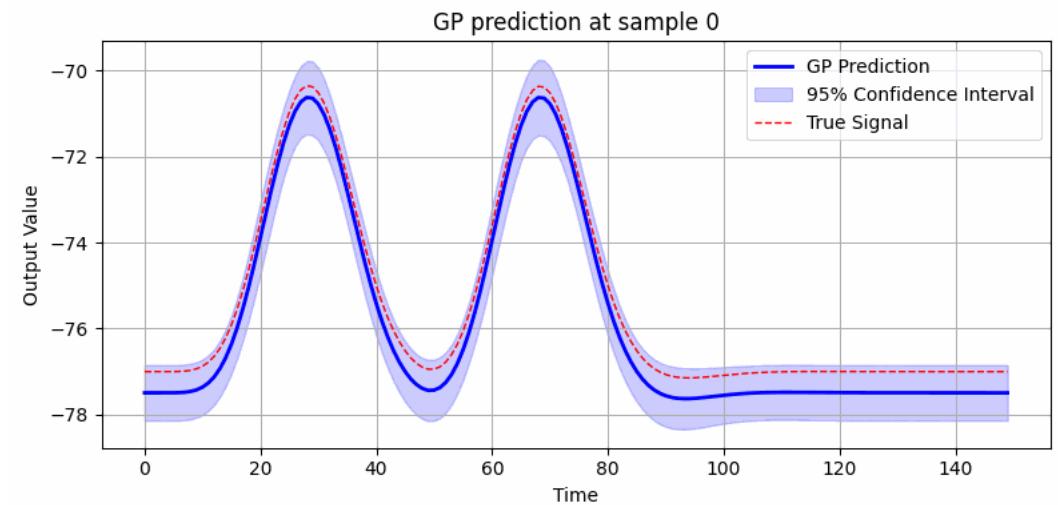
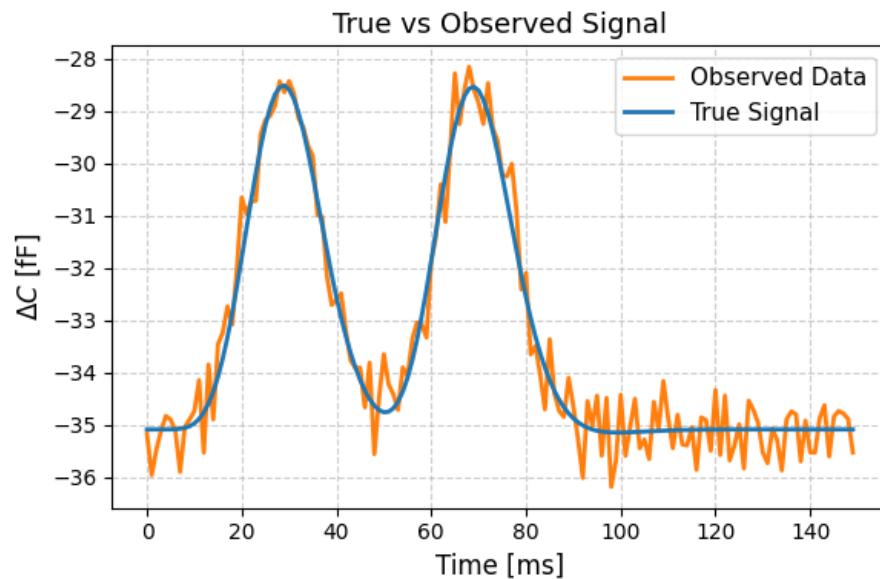
where

$$r^2(\mathbf{x}, \mathbf{x}') = \sum_{d=1}^D \frac{(x_d - x'_d)^2}{\ell_d^2}$$

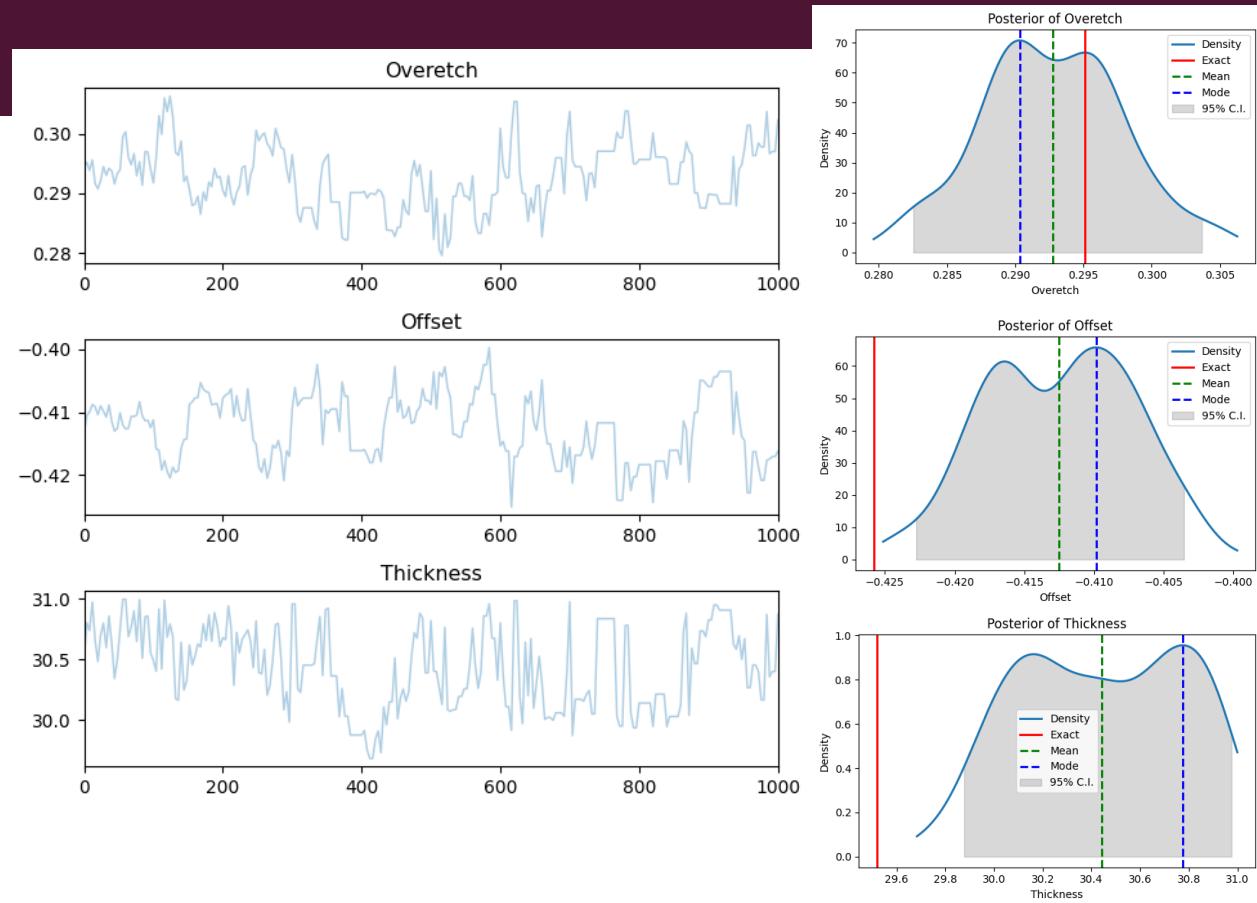
ACTIVE SAMPLING STRATEGY

- Use a Delayed Acceptance Algorithm for solving the inverse problem.
- At coarse level we use the GP - likelihood, adjusted by the covariance.
$$\mathcal{L}_{\text{GP}}(y \mid \mathbf{x}) = \mathcal{N}(y \mid \mu_{\text{GP}}(\mathbf{x}), \Sigma_{\text{obs}} + \Sigma_{\text{GP}}(\mathbf{x}))$$
- At fine level, we use the FOM.
$$\mathcal{L}_{\text{FOM}}(y \mid \mathbf{x}) = \mathcal{N}(y \mid \mathcal{F}(\mathbf{x}), \Sigma_{\text{obs}})$$
- Every 5 steps, we fit the GP introducing New Point.
- When the GP standard deviation becomes negligible, we stop training and switch to only surrogate

ONLINE SAMPLING



GP - MCMC



CONCLUSION

- The active sampling allows to reduce by a factor of 4 the total number of samples need to train the GP.
- Numerical Issues in inverse correction reduce the effectiveness of the method, for which a regularization is introduced, limiting the accuracy of the method.
- Preconditioning techniques should be investigated.
- In general, the method allows to drastically reduce the computational times needed for MCMC with good accuracy.

BIBLIOGRAPHY

- Marinescu, Marius. "Explaining and Connecting Kriging with Gaussian Process Regression." *arXiv preprint arXiv:2408.02331*(2024).
- Helin, Tasio, et al. "Introduction To Gaussian Process Regression In Bayesian Inverse Problems, With New ResultsOn Experimental Design For Weighted Error Measures." *arXiv preprint arXiv:2302.04518* (2023).
- Zachei, Filippo, et al. "Neural networks based surrogate modeling for efficient uncertainty quantification and calibration of MEMS accelerometers." *International Journal of Non-Linear Mechanics* 167 (2024): 104902.
- [2] Benjamin Peherstorfer, Karen Willcox, and Max Gunzburger. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *Siam Review*,60(3):550–591, 2018.