Estimation of an improved surrogate model in uncertainty quantification by neural networks

Sebastian Kersting and Michael Kohler
CRC 805, Technische Universität Darmstadt, email: kersting@mathematik.tu-darmstadt.de, kohler@mathematik@tu-darmstadt.de

1 Introduction

Our main goal is to estimate the uncertainty of the maximal oscillation amplitude of a lateral vibration attenuation system with piezo-elastic supports.

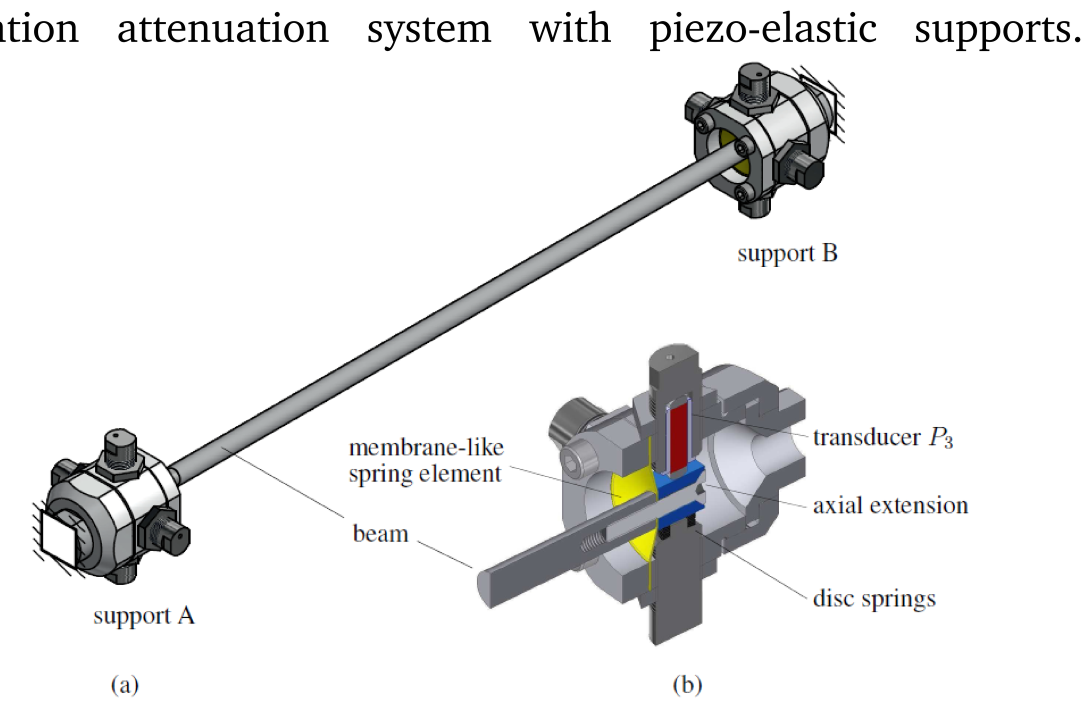
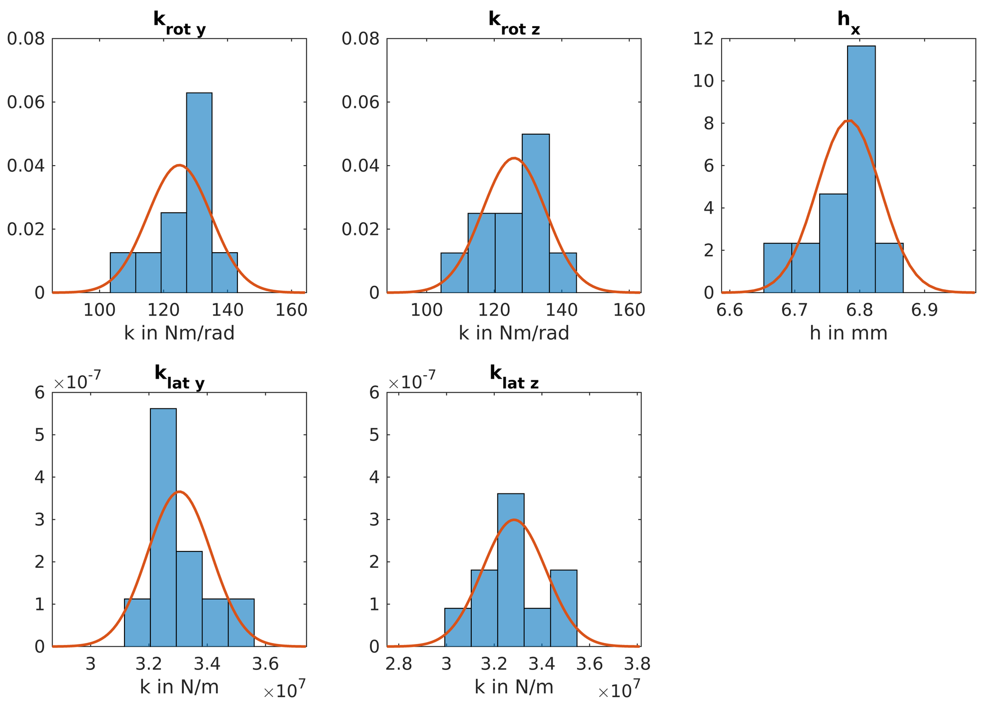


Figure 1: (a) CAD model of the beam with piezo-elastic supports, (b) Sectional view of piezo-elastic support B

We assume the stochastic uncertainty in the Input data is known, e.g. by given density.



We assume a computer experiment of the underlying physical model is given in conjunction with a small sample of experimental data. Based on these input we want to quantify the uncertainty by estimating the density of the maximal oscillation amplitude.

2 Mathematical description of the problem

- Let X be a $\mathbb{R}^d$-dimensional random variable describing the initial state, Y the dependent $\mathbb{R}$-valued random variable which density we want to estimate and $m: \mathbb{R}^d \to \mathbb{R}$ be a function.
- $m(X)$ is an (imperfect) approximation of Y.
- $(X_1, Y_1), \dots, (X_n, Y_n)$ is a small sample of experimental data.
- $(X_{n+1}, m(X_{n+1})), \dots, (X_{n+L_n}, m(X_{n+L_n}))$ is data generated by computer a experiment.
- $X_{n+L_n+1}, \dots, X_{n+L_n+N_n}$ are additional input values.

3 Generalized hierarchical interaction model

A function $m: \mathbb{R}^d \to \mathbb{R}$:

- satisfies a generalized hierarchical interaction model of order d^* and level 0, if there exist $a_1, \dots, a_{d^*} \in \mathbb{R}^d$ and $f: \mathbb{R}^{d^*} \to \mathbb{R}$ such that
$$m(x) = f(a_1^T x, \dots, a_{d^*}^T x) \quad \text{for all } x \in \mathbb{R}^d.$$
- satisfies a generalized hierarchical interaction model of order d^* and level $l+1$, if there exist $K \in \mathbb{N}$, $g_k: \mathbb{R}^{d^*} \to \mathbb{R}$ $(k = 1, \dots, K)$ and $f_{1,k}, \dots, f_{d^*,k}: \mathbb{R}^d \to \mathbb{R}$ $(k = 1, \dots, K)$ such that $f_{1,k}, \dots, f_{d^*,k}$ $(k = 1, \dots, K)$ satisfy a generalized hierarchical interaction model of order d^* and level l and
$$m(x) = \sum_{k=1}^{K} g_k(f_{1,k}(x), \dots, f_{d^*,k}(x)) \quad \text{for all } x \in \mathbb{R}^d.$$

The generalized hierarchical interaction model is (p, C)-smooth, if all functions occurring in its definition are (p, C)-smooth.

4 Definition of the estimate

- Denote the set of all functions $f: \mathbb{R}^d \to \mathbb{R}$ that satisfy
$$f(x) = \sum_{i=1}^{M} \mu_i \cdot \sigma\left(\sum_{j=1}^{4d^*} \lambda_{i,j} \cdot \sigma\left(\sum_{v=i}^{d} \theta_{i,j,v} \cdot x^{(v)} + \theta_{i,j,0}\right) + \lambda_{i,0}\right) + \mu_0$$
$(x \in \mathbb{R}^d)$ where the factors $\mu_i, \lambda_{i,j}, \theta_{i,j,v} \in \mathbb{R}$ are bounded in absolute value by some $B_n > 0$ by $\mathcal{F}^{(\text{neural networks})}_{M,d,d^*,B_n}$. As function $\sigma: \mathbb{R} \to \mathbb{R}$ we use the logistic squasher.
- For $l = 0$: $\mathcal{H}^{(0)}_{K,M,d,d^*,B_n} = \mathcal{F}^{(\text{neural networks})}_{M,d,d^*,B_n}$.
- For $l > 0$:
$$\mathcal{H}^{(l)}_{K,M,d,d^*,B_n} = \left\{h: \mathbb{R}^d \to \mathbb{R},\ h(x) = \sum_{k=1}^{K} g_k(f_{1,k}(x), \dots, f_{d^*,k}(x)) \quad (x \in \mathbb{R}^d) \text{ for some } g_k \in \mathcal{F}^{(\text{neural networks})}_{M,d^*,d^*,B_n} \text{ and } f_{j,k} \in \mathcal{H}^{(l-1)}_{K,M,d,d^*,B_n}\right\}$$
- Define a surrogate least squares estimate
$$\tilde{m}_{L_n}(\cdot) = \arg\min_{h \in \mathcal{H}^{(l)}_{K,M_{L_n},d,d^*,B_n}} \left(\frac{1}{L_n} \sum_{i=n+1}^{n+L_n} |h(X_i) - m(X_i)|^2\right)$$
and $m_{L_n} = T_\beta(\tilde{m}_{L_n})$where $T_\beta(z) = \text{sign}(z) \cdot \min\{|z|, \beta\}$ for some $\beta > 0$.
- Set $\hat{\epsilon}_i = Y_i - m_{L_n}(X_i)$, for $i = 1, \dots, n$.
- Define a weighted least squares residuals estimate by
$$\tilde{m}_n^{\hat{\epsilon}}(\cdot) = \arg\min_{h \in \mathcal{H}^{(l)}_{K,M_n,d,d^*,B_n}} \left(\frac{w^{(n)}}{n} \sum_{i=1}^{n} (\hat{\epsilon}_i - h(X_i))^2 + \frac{1 - w^{(n)}}{N_n} \sum_{i=1}^{N_n} (0 - h(X_{n+L_n+i}))^2\right)$$
and $\hat{m}_n^{\hat{\epsilon}} = T_{c_1 \cdot \alpha_n}(\tilde{m}_n^{\hat{\epsilon}})$.
- Set
$$\hat{m}_n(x) = m_{L_n}(x) + \hat{m}_n^{\hat{\epsilon}}(x) \quad (x \in \mathbb{R}^d).$$
- The density g of Y will be estimated by applying a kernel density estimate to a sample of $\hat{m}_n(X)$. Therefore we choose a kernel $K: \mathbb{R} \to \mathbb{R}$ and a bandwidth $h_{N_n} > 0$ and define
$$\hat{g}_{N_n}(y) = \frac{1}{N_n \cdot h_{N_n}} \cdot \sum_{i=1}^{N_n} K\left(\frac{y - \hat{m}_n(X_{n+L_n+i})}{h_{N_n}}\right).$$

5 Main Result

Theorem 1 *Let $d, n, L_n \in \mathbb{N}$ with $2 \le n \le L_n \le N_n$, with $L_n \le n^{c_2}$. Let (X, Y), (X_1, Y_1), ... be independent and identically distributed $\mathbb{R}^d \times \mathbb{R}$–valued random variables with $\mathbf{E}\{|Y|\} < \infty$ and with $supp(X)$ bounded. Let $m^*(\cdot) = \mathbf{E}\{Y|X = \cdot\}$ be the regression function. Let $m: \mathbb{R}^d \to \mathbb{R}$ be a (p, C)-smooth measurable function, which satisfies a (p, C)-smooth hierarchical interaction model of order d^* and finite level l with $p = q + s$ for some $q \in \mathbb{N}_0$ and $s \in (0, 1]$. Furthermore assume that in all partial derivates of order less than or equal to q of the functions of this hierarchical interaction model g_k are bounded, i.e., assume that each such function f satisfies*
$$\max_{\substack{j_1,\dots,j_d \in \{0,1,\dots,d\} \\ j_1+\dots+j_d \le q}} \left\| \frac{\partial^{j_1+\dots+j_d} f}{\partial^{j_1} x^{(1)} \cdots \partial^{j_d} x^{(d)}} \right\|_\infty \le c_3, \qquad (1)$$
and let all functions g_k be Lipschitz continuous with Lipschitz constant $L > 0$. Assume for some $1 \le \beta \le n + L_n$
$$|m(x)| < \beta.$$
Let $\alpha_n > \alpha_n^* \ge 0$ and assume
$$\mathbf{E}\left\{|Y - m^*(X)|^2\right\} \le (\alpha_n^*)^2 \quad \text{and} \quad \mathbf{E}\left\{|Y - m^*(X)|^3\right\} \le (\alpha_n^*)^3,$$
that there exists $K, \sigma_0 > 0$ such that
$$K^2 \cdot \left(\mathbf{E}\left\{\exp\left(\frac{(Y - m^*(X))^2}{\alpha_n \cdot K}\right) \middle| X\right\} - 1\right) \le \sigma_0 \quad a.s.,$$
that the regression function $\mathbf{E}\{\frac{1}{\alpha_n}(Y - m(X))|X = x\} = \frac{1}{\alpha_n}(m^* - m)(x)$ satisfies a (p, C)-smooth hierarchical interaction model of order d^* and finite level l. Furthermore assume that all partial derivates of order less than or equal to q of the functions of this hierarchical interaction model g_k are bounded as analogue to (1) and let all functions g_k be Lipschitz continuous with Lipschitz constant $L > 0$. Assume
$$\sup_{x \in \mathbb{R}^d} \left| \frac{1}{\alpha_n} \cdot (m^* - m)(x) \right| \le c_5$$
and
$$\left(\log(L_n)^3 \cdot L_n^{-\frac{2p}{2p+d^*}}\right)^{1/3} \le \alpha_n \le c_6.$$
Let $m_{L_n} \in \mathcal{H}^{(l)}_{K,M_{L_n},d,d^*,B_n}$ and $\hat{m}_n^{\hat{\epsilon}} \in \mathcal{H}^{(l)}_{K,M_n,d,d^*,B_n}$ where $\mathcal{H}^{(l)}_{K,M_{L_n},d,d^*,B_n}$ and $\mathcal{H}^{(l)}_{K,M_n,d,d^*,B_n}$ are sets of hierarchical neural networks with K, d, d^* as in the definition of m^*, $M_{L_n} = \left\lceil c_7 \cdot L_n^{\frac{d^*}{2p+d^*}} \right\rceil$, $M_n = \left\lceil c_8 \cdot n^{\frac{d^*}{2p+d^*}} \right\rceil$ and $B_n = n^{c_9}$. Then there exists constants $c_{10}, \dots, c_{15} \in \mathbb{R}_+$ such that
$$\mathbf{E}\left\{|Y - \hat{m}_n(X)|^2\right\} \le c_{10} \cdot (\alpha_n^*)^2 + c_{11} \cdot \alpha_n^2 \cdot \log(n)^3 \cdot n^{-\frac{2p}{2p+d^*}} + c_{12} \cdot \log(L_n)^3 \cdot L_n^{-\frac{2p}{2p+d^*}} + c_{13} \cdot (1 - w^{(n)}) \cdot \alpha_n^2 + \frac{c_{14} \cdot \alpha_n^2}{n} + \frac{c_{15}}{L_n}.$$
In particular, in case $w^{(n)} = 1$ we get
$$\mathbf{E}\left\{|Y - \hat{m}_n(X)|^2\right\} \le c_{16} \cdot \max\left\{(\alpha_n^*)^2, \alpha_n^2 \cdot \log(n)^3 \cdot n^{-\frac{2p}{2p+d^*}}, \log(L_n)^3 \cdot L_n^{-\frac{2p}{2p+d^*}}\right\}$$
for some $c_{16} \in \mathbb{R}^d$.

Corollary 1 *Assume that the density g of Y is (r, C)-smooth for some $r \in (0, 1]$ and that its support is compact. Let $K: \mathbb{R} \to \mathbb{R}$ be a symmetric and bounded density which decreases monotonically on $\mathbb{R}_+$. Assume that the assumptions of Theorem 1 are satisfied, and that, in addition,*
$$\max\left\{(\alpha_n^*)^2, \log(L_n)^3 \cdot L_n^{-\frac{2p}{2p+d^*}}\right\} \le \alpha_n^2 \cdot \log(n)^3 \cdot n^{-\frac{2p}{2p+d^*}}$$
holds. Set
$$h_{N_n} = c_{17} \cdot \left(\alpha_n \cdot \log(n)^{3/2} \cdot n^{-\frac{2p}{2p+d^*}}\right)^{\frac{1}{r+1}}$$
and assume
$$N_n \ge \left(\frac{n^{\frac{2p}{2p+d^*}}}{\alpha_n \cdot \log(n)^3}\right)^{\frac{2r+1}{r+1}}$$
Then we have
$$\mathbf{E} \int_{\mathbb{R}} |\hat{g}_{N_n}(y) - g(y)| \, dy \le c_{18} \cdot \left(\alpha_n \cdot \log(n)^{3/2} \cdot n^{-\frac{p}{2p+d^*}}\right)^{\frac{r}{r+1}}$$
for some $c_{18} \in \mathbb{R}_+$.

6 Application to simulated data

Set
$$Y = m^*(X) + \sigma^* \cdot 9.11 \cdot \epsilon$$
with $X \sim U([0, 1]^7)$, $\sigma > 0$ and $\epsilon \sim U([-1, 1])$ such that X and ϵ are independent. Set
$$m^*(x) = \cot\left(\frac{\pi}{1 + \exp\left(x_1^2 + 2 \cdot x_2 + sin(6 \cdot x_4^3) - 3\right)}\right).$$
The stochastic model is defined by
$$m(x) = m^*(x) + \sigma_m \cdot 2.25 \quad (x \in \mathbb{R}^7)$$

L_2 error of the improved and the surrogate estimate for different values of σ^* and σ_m for sample sizes $n = 10$, $L_n = 200$, $N_n = 200$:

σ^*	5%			20 %		
σ_m	0.5	1	5	0.5	1	5
$\hat{m}_n$	5.86	6.14	5.77	7.10	6.35	6.76
m_{L_n}	6.71	10.76	137.28	7.34	11.45	137.57

L_2 error of a neural network trained on n realizations of (X, Y):

σ^*	5%			20 %		
n	10	100	200	10	100	200
L_2 error	568.57	10.68	5.12	567.44	13.79	6.74