

Estimation of Uncertainty in the Lateral Vibration Attenuation of a Beam with Piezo-Elastic Supports by Neural Networks

Benedict Götz^{1,a}, and Sebastian Kersting^{2,b}

¹Fachgebiet Systemzuverlässigkeit, Adaptronik und Maschinenakustik SAM, Technische Universität Darmstadt, Magdalenenstr. 4, 64289 Darmstadt, Germany

²Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany

^a goetz@szm.tu-darmstadt.de, ^bkersting@mathematik.tu-darmstadt.de

Keywords: Density estimation, surrogate model, neural network, imperfect model, uncertainty quantification.

Abstract. Quantification of uncertainty in technical systems is often based on surrogate models of corresponding simulation models. Usually, the underlying simulation model does not describe the reality perfectly, and consequently the surrogate model will be imperfect. In this article we propose an improved surrogate model of the vibration attenuation of a beam with shunted piezoelectric transducers. Therefore, experimentally observed and simulated variations in the vibration attenuation are combined in the model estimation process, by using multi-layer feedforward neural networks. Based on this improved surrogate model, we construct a density estimate of the maximal amplitude in the vibration attenuation. The density estimate is used to analyze the uncertainty in the vibration attenuation, resulting from manufacturing variations.

Introduction

In this article we introduce a new method for uncertainty quantification in complex technical systems. Our main goal is to estimate the uncertainty in the maximal vibration amplitude of a beam system with piezo-elastic supports described in Figure 1. This system consists of a beam with circular cross-

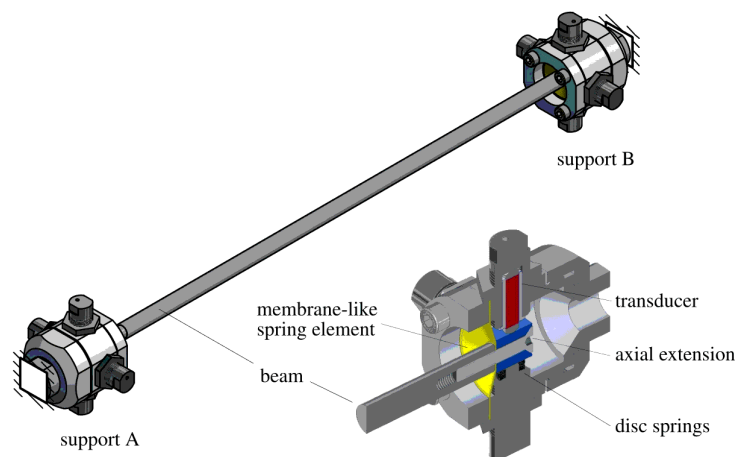


Fig. 1: A CAD model of the lateral vibration attenuation system with piezo-elastic supports and a sectional view of one of the piezo-elastic supports, cf. [3].

section embedded in two piezo-elastic supports A and B where support A is used for lateral beam vibration excitation and support B is used for lateral beam vibration attenuation, as proposed in [3]. The two piezo-elastic supports A and B are located at the beam's end and each consist of one elastic membrane-like spring element made of spring steel, two piezoelectric stack transducers arranged orthogonally to each other and mechanically prestressed with disc springs as well as the relatively

stiff axial extension made of hardened steel that connects the piezoelectric transducers with the beam. For vibration attenuation in support B, optimally tuned electrical shunt circuits are connected to the piezoelectric transducers.

If we produce and assemble the piezo–elastic support several times, the system properties will differ due to manufacturing and assembly variations. Consequently the outcome of an experiment with such a system will vary. E.g., building such systems ten times and measuring the maximal attenuated vibration amplitude in an experiment with each of the built systems, we got the following ten values:

$$\begin{aligned} y_1 = 14.50, y_2 = 14.17, y_3 = 14.37, y_4 = 14.16, y_5 = 14.28, y_6 = 13.51, \\ y_7 = 14.73, y_8 = 13.21, y_9 = 13.05, y_{10} = 16.26. \end{aligned} \quad (1)$$

We assume in the sequel that y_1, \dots, y_{10} are independent realizations of a real-valued random variable Y , and in order to get information about the distribution of Y we try to estimate the density $g : \mathbb{R} \rightarrow \mathbb{R}$ of Y with respect to the Lebesgue measure (which we assume to exist).

The classical statistical approach of doing this is to assume that Y is, e.g., normally distributed, to estimate its mean and its variance by maximum likelihood and to use the density of the corresponding normal distribution as an estimate of the density of Y . For the data in (1) this results in the black solid curve in Figure 2. However, the maximum vibration amplitudes represent extreme values of the lateral beam vibration transfer behavior. According to [2], the distribution of extreme values is characterized by a non-symmetric distribution about the most likely value. Thus, this approach does not seem promising.

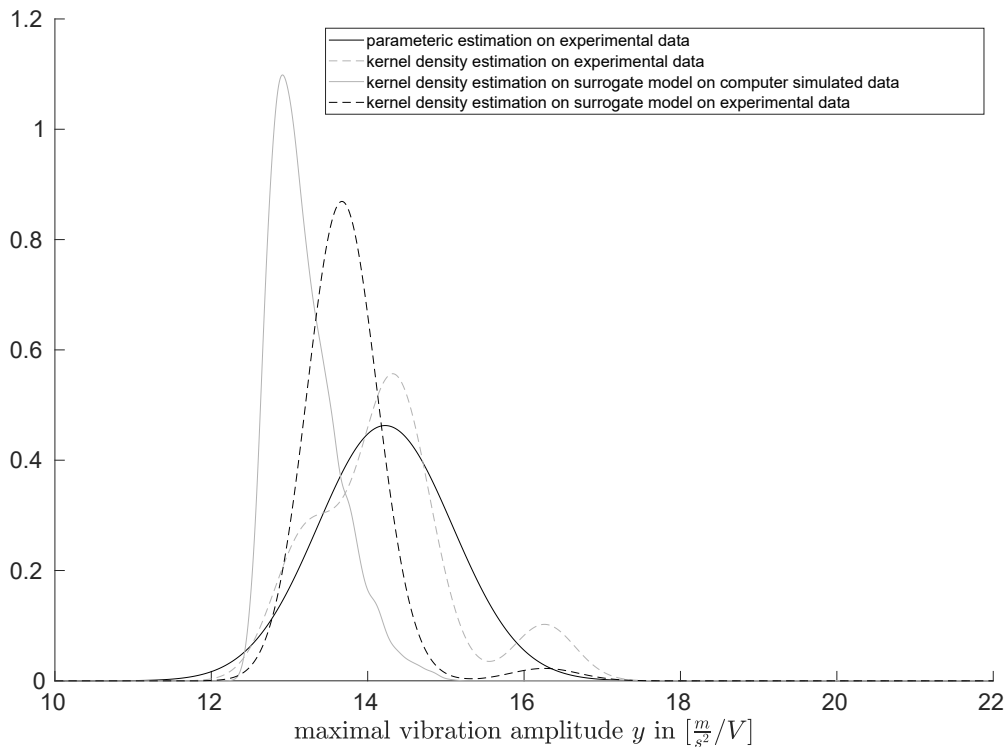


Fig. 2: A parametric (black solid line), a nonparametric (gray dashed line) estimate of the density of the data (1). A surrogate estimate on experimental data (5) (black dashed line) and a surrogate estimate on computer simulated data (8) (gray solid line).

The standard approach in modern statistics would be to use a nonparametric estimate of the density of Y , e.g. the classical kernel density estimate of [8] and [7]

$$\hat{g}_{Y,n}(y) = \frac{1}{n \cdot h_n} \cdot \sum_{i=1}^n K\left(\frac{y - Y_i}{h_n}\right), \quad (2)$$

which we apply in the above formula to random variables Y_1, \dots, Y_n which are independent identically distributed as Y . Here $K : \mathbb{R}^d \rightarrow \mathbb{R}$ (so-called kernel, which is assumed to be a density) and $h_n > 0$ (so-called bandwidth) are parameters of the estimate. E.g., computing this kernel density estimate with the routine *ksdensity()* in *MATLAB* results in the gray dashed curve in Figure 2.

The obvious drawback of the first approach is that the error of this parametric estimate might be rather large in case that the true density of Y is not the density of a normal distribution, in particular if it cannot be approximated well by any such density. However, due to the small sample size in this example it is not clear that the second approach, i.e., the nonparametric density estimate, yields an estimate which is better than the parametric estimate. So in general neither of these two approaches will lead to satisfying results.

Unfortunately, due to cost and time limitations, it is not really possible to increase the sample size 10 of the data (1) in such a way that a nonparametric estimate seems promising, since experiments with the above system (in particular the construction and replacement of the membrane-like spring elements) are extremely time consuming. What we do instead in the sequel is to use some knowledge outside of the data (e.g., knowledge from engineering science about attenuation systems) in order to improve our estimation.

Often this is done in the framework of Bayesian statistics, where some kind of a priori distribution describing the system under consideration is assumed to be given, and under the assumption that this is indeed true, estimates are constructed which achieve good results even for very small sample sizes. However, this is an example of the saying "We buy information with assumptions" [1], which of course might lead to wrong informations in the case of wrong assumptions. And since it is not obvious how to transform the knowledge in engineering science into assumptions about an a priori distribution, we will not use this approach.

Instead, we will use the following knowledge in engineering science in order to construct an improved estimate: It is known that five parameters of the membrane in the attenuation system vary during the construction of the attenuation system and influence the maximal vibration amplitude: the lateral stiffness in direction of y ($k_{lat,y}$) and in direction of z ($k_{lat,z}$), the rotatory stiffness in direction of y ($k_{rot,y}$) and in direction of z ($k_{rot,z}$), and the height of the membrane (h_x). For given values of these five parameters it is possible to compute in a physical model of the attenuation system the corresponding maximal vibration amplitude. In order to generate values of these five parameters we need to determine their distributions. Therefore we measured the corresponding parameters for the ten built systems. As a result we got the data in Table 1. The four discrete stiffness properties as well as the height are assumed to be normally distributed and besides we assume that they are independent. Consequently we use the 10 data points for each of the five parameters in order to estimate parametrically a normal distribution by a maximum likelihood estimate for each of these parameters, cf., Figure 3. With the physical model and the now known distribution of the 5 dimensional random vector X we can compute the outcome Y of an experiment with our technical system by a computer program $m : \mathbb{R}^5 \rightarrow \mathbb{R}$. We assume that the computer program is a good approximation of the values we get for the maximal vibration amplitude in experiments.

In this stochastic model of our attenuation system we can generate independent data $X_{n+1}, \dots, X_{n+L_n}$, compute $m(X_{n+1}), \dots, m(X_{n+L_n})$ and define a kernel density estimate by

$$\hat{g}_{L_n}(y) = \frac{1}{L_n \cdot h_{L_n}} \cdot \sum_{i=1}^{L_n} K\left(\frac{y - m(Y_{n+i})}{h_{L_n}}\right).$$

Table 1: Measured data for the ten built systems. The values of $k_{rot,y}$ and $k_{rot,z}$ are given in $[Nm/\text{rad}]$, the values of $k_{lat,y}$ and $k_{lat,z}$ are given in $[N/m]$, the values of h_x are given in $[m]$ and the values of y are given in $[\frac{m}{s^2}/v]$.

	1	2	3	4	5	6	7	8	9	10
$k_{rot,y} \times 10^2$	1.31	1.34	1.31	1.23	1.14	1.29	1.35	1.28	1.04	1.20
$k_{rot,z} \times 10^2$	1.31	1.28	1.43	1.25	1.30	1.34	1.22	1.16	1.18	1.11
$k_{lat,y} \times 10^7$	3.27	3.28	3.35	3.29	3.22	3.26	3.19	3.54	3.21	3.42
$k_{lat,z} \times 10^7$	3.07	3.22	3.29	3.25	3.30	3.18	3.16	3.51	3.37	3.44
$h_x \times 10^{-4}$	6.79	6.77	6.82	6.80	6.79	6.76	6.81	6.74	6.68	6.84
$y \times 10^1$	1.45	1.42	1.44	1.42	1.43	1.35	1.47	1.32	1.31	1.63

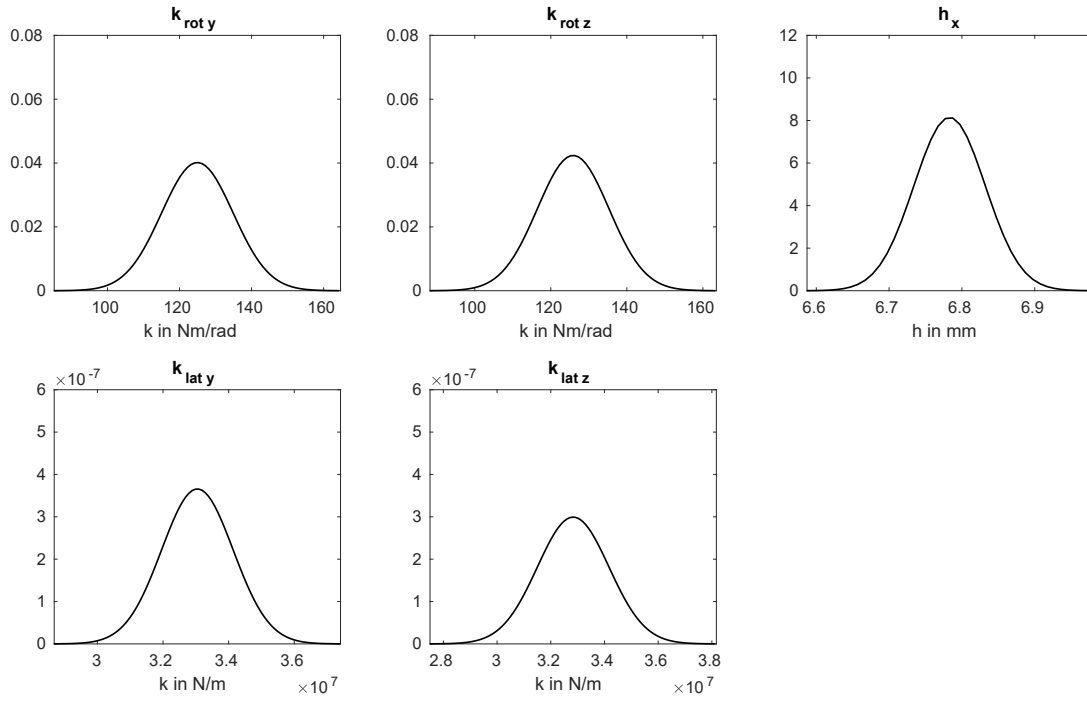


Fig. 3: Parametric estimates for the densities of the five parameters of our simulation model.

However, the evaluation of the computer program for our technical system will often be rather time consuming and consequently L_n (although much larger than n) might not be really large. One possibility to circumvent this problem is to define an estimate of g on the basis of the data

$$(X_{n+1}, m(X_{n+1})), \dots, (X_{n+L_n}, m(X_{n+L_n})), X_{n+L_n+1}, \dots, X_{n+L_n+N_n} \quad (3)$$

by estimating in a first step a surrogate

$$m_{(X, m(X)), L_n}(\cdot) = m_{(X, m(X)), L_n}(\cdot, (X_{n+1}, m(X_{n+1})), \dots, (X_{n+L_n}, m(X_{n+L_n}))) : \mathbb{R}^d \rightarrow \mathbb{R} \quad (4)$$

of m and by defining in a second step the corresponding surrogate density estimate via

$$\hat{g}_{(X, m(X)), L_n}(y) = \frac{1}{N_n \cdot h_{N_n}} \cdot \sum_{i=1}^{N_n} K\left(\frac{y - m_{(X, m(X)), L_n}(X_{n+L_n+i})}{h_{N_n}}\right). \quad (5)$$

Computing such an surrogate density estimate results in the yellow line in Figure 2. Alternatively, one can also ignore the simulation model completely, and can use instead the data

$$(X_1, Y_1), \dots, (X_n, Y_n), X_{n+L_n+1}, \dots, X_{n+L_n+N_n} \quad (6)$$

in order to construct an estimate

$$m_{(X,Y),n}(\cdot) = m_{(X,Y),n}(\cdot, (X_1, Y_1), \dots, (X_n, Y_n)) : \mathbb{R}^d \rightarrow \mathbb{R} \quad (7)$$

of $m^*(x) = \mathbf{E}\{Y|X = x\}$, and can define the corresponding surrogate density estimate by

$$\hat{g}_{(X,Y),n}(y) = \frac{1}{N_n \cdot h_{N_n}} \cdot \sum_{i=1}^{N_n} K \left(\frac{y - m_{(X,Y),n}(X_{n+L_n+i})}{h_{N_n}} \right). \quad (8)$$

The main question which we want to investigate by simulations in this paper is whether there exist situations in which suitably defined estimates based on the complete data

$$(X_1, Y_1), \dots, (X_n, Y_n), (X_{n+1}, m(X_{n+1})), \dots, (X_{n+L_n}, m(X_{n+L_n})), \\ X_{n+L_n+1}, \dots, X_{n+L_n+N_n} \quad (9)$$

(where $n, L_n, N_n \in \mathbb{N}$) achieve simultaneously better rate of convergence results than the estimates (2), (5) and (8).

A Method for Estimating Uncertainty by Improving a Surrogate Model with Real Data

In the following section we introduce a method which combines experimental and computer model data to generate an estimate of the probability density function $g: \mathbb{R} \rightarrow \mathbb{R}$ of Y .

Definition of a class of neural networks In order to construct such an estimate, we proceed as follows: Let $\sigma: \mathbb{R} \rightarrow \mathbb{R}$ be the so-called logistic squasher $\sigma(x) = 1/(1 + \exp(-x))$ ($x \in \mathbb{R}$).

For $M \in \mathbb{N}$, $d \in \mathbb{N}$ and $d^* \in \{0, \dots, d\}$, we denote the set of all functions $f: \mathbb{R}^d \rightarrow \mathbb{R}$ that satisfy

$$f(x) = \sum_{i=1}^M \mu_i \cdot \sigma \left(\sum_{j=1}^{4d^*} \lambda_{i,j} \cdot \sigma \left(\sum_{v=1}^d \theta_{i,j,v} \cdot x^{(v)} + \theta_{i,j,0} \right) + \lambda_{i,0} \right) + \mu_0$$

($x \in \mathbb{R}^d$) for some $\mu_i, \lambda_{i,j}, \theta_{i,j,v} \in \mathbb{R}$, by $\mathcal{F}_{M,d,d^*}^{(\text{neural networks})}$.

We will use the following recursively defined classes of neural networks (with parameters $K, M, d, d^* \in \mathbb{N}$): For $l = 0$, we define our space of hierarchical neural networks by

$$\mathcal{H}_{K,M,d,d^*}^{(0)} = \mathcal{F}_{M,d,d^*}^{(\text{neural networks})}.$$

For $l > 0$, we define recursively

$$\mathcal{H}_{K,M,d,d^*}^{(l)} = \left\{ h: \mathbb{R}^d \rightarrow \mathbb{R}, h(x) = \sum_{k=1}^K g_k(f_{1,k}(x), \dots, f_{d^*,k}(x)) \quad (x \in \mathbb{R}^d) \right. \\ \left. \text{for some } g_k \in \mathcal{F}_{M,d^*,d^*}^{(\text{neural networks})} \text{ and } f_{j,k} \in \mathcal{H}_{K,M,d,d^*}^{(l-1)} \right\}. \quad (10)$$

These classes of neural networks were introduced in Kohler and Krzyżak (2017) as an approximation for so-called hierarchical interaction models. We do not explain the advantages of hierarchical interaction models at this point, but we want to remark that it is a realistic assumption in connection with complex technical systems which are built in a modular way.

Definition of the estimate Given the data from Table 1 and a computer model of the technical system we want to estimate the density g of Y .

We start by estimating the distribution of our independent variables X with a maximum likelihood estimate, where we assume that the underlying distribution is known. With the estimated distribution of X we generate a samples of size L_n and $N_{2,n}$. We use the first sample to generate the data set (3). The second one is necessary in the density estimation later on.

Next we define a surrogate estimate

$$m_{L_n}(\cdot) = m_{L_n}(\cdot, (X_{n+1}, m(X_{n+1})), \dots, (X_{n+L_n}, m(X_{n+L_n}))) : \mathbb{R}^d \rightarrow \mathbb{R} \quad (11)$$

of the function m . For this we use a least squares estimate defined by

$$m_{L_n}(\cdot) = \arg \min_{h \in \mathcal{H}_{K_1, M_{1,n}, d, d^*}^{(l)}} \left(\frac{1}{L_n} \sum_{i=n+1}^{n+L_n} |h(X_i) - m(X_i)|^2 \right), \quad (12)$$

where $K_1, M_{1,n}, d^* \in \mathbb{N}$ are parameters of the estimate. For simplicity we assume here and in the sequel that the minimum above indeed exists and to compute an approximation of the minimum we use the Levenberg-Marquardt algorithm, cf. [5] or [6]. In the implementation of the surrogate estimate we use the *MATLAB* routine *lsqnonlin()* which implements this algorithm. To choose suitable parameters $K_1, M_{1,n}$ and d^* we use the splitting of the sample technique, where we calculate the minimum on $\frac{2}{3} \cdot \lceil L_n \rceil$ training data and evaluate the L_2 error on the remaining $L_n - \frac{2}{3} \cdot \lceil L_n \rceil$ test data. The parameter combination with the smallest L_2 risk is selected, where we choose parameter from the sets $l \in \{0, 1, 2\}$, $K_1 \in \{1, 2\}$, $d^* \in \{1, \dots, d\}$ and $M_{1,n} \in \{1, \dots, 5, 6, 16, \dots, 46\}$.

Next we define an estimate of $m^* - m_{L_n}$ on the basis of the residuals

$$\hat{\epsilon}_i = Y_i - m_{L_n}(X_i) \quad (i = 1, \dots, n). \quad (13)$$

Therefore we define

$$\hat{m}_n^{\hat{\epsilon}}(\cdot) = \arg \min_{h \in \mathcal{H}_{K_2, M_{2,n}, d, d^*}^{(l)}} \left(\frac{w^{(n)}}{n} \sum_{i=1}^n (\hat{\epsilon}_i - h(X_i))^2 + \frac{1 - w^{(n)}}{N_{1,n}} \sum_{i=1}^{N_{1,n}} (0 - h(X_{n+L_n+i}))^2 \right) \quad (14)$$

for some weight $w^{(n)} \in [0, 1]$ and parameters $K_2, M_{2,n}, d^* \in \mathbb{N}$. Again we use the Levenberg-Marquardt algorithm to compute an approximation of the solution of the above minimization problem. Due to the small sample size of the data set $(X_1, Y_1), \dots, (X_n, Y_n)$ we reduce the parameter sets to $l \in \{0\}$, $K_2 \in \{1\}$, $d^* \in \{1, 2, 4\}$, $M_{2,n} \in \{1, 3, 5\}$ and the additional weighting parameter w is chosen from $\{0, 0.25, \dots, 1\}$ and use a 5-fold cross validation to evaluate the parameter combination with the smallest L_2 risk.

We define our final surrogate model $(X, \hat{m}_n(X))$ for (X, Y) by

$$\hat{m}_n(x) = m_{L_n}(x) + \hat{m}_n^{\hat{\epsilon}}(x) \quad (x \in \mathbb{R}^d). \quad (15)$$

To estimate the density g of Y we apply a kernel density estimate to the sample

$$\hat{m}_n(X_{n+L_n+N_{1,n}+1}), \dots, \hat{m}_n(X_{n+L_n+N_{1,n}+N_{2,n}}).$$

Therefore we choose a kernel $K : \mathbb{R} \rightarrow \mathbb{R}$ and a bandwidth $h_{N_{2,n}} > 0$ and set

$$\hat{g}_{N_{2,n}}(y) = \frac{1}{N_{2,n} \cdot h_{N_{2,n}}} \cdot \sum_{i=1}^{N_{2,n}} K \left(\frac{y - \hat{m}_n(X_{n+L_n+N_{1,n}+i})}{h_{N_{2,n}}} \right), \quad (16)$$

where we use the *MATLAB* routine *ksdensity()*.

Application to Simulated and Real Data

To be able to rate the performance of our newly proposed estimate we compare it to other classical approaches in a simulated environment. Finally we will apply the new method to the data from our lateral vibration attenuation system.

We compare our estimate (est. 4) with three other density estimates. The first one (est. 1) is a standard kernel density estimate applied to a sample of size n of Y , cf. (2). The estimates 2 and 3 are surrogate density estimates where the kernel density estimate of *MATLAB* is applied to a sample of size $N_{2,n}$ of the surrogate model. For the second estimate (est. 2) the surrogate model is chosen as a neural network trained on L_n realizations of $(X, m(X))$, cf. (5). For the third estimate (est. 3) the surrogate model is chosen as a neural network trained on n realizations of (X, Y) , cf. (8).

In all simulations we neglect the parametric density estimation of the input variable X and assume a perfect approximation of the density is available. We choose X as a uniformly distributed random vector on $[0, 1]^d$. Furthermore we choose a random error ϵ which is independent from X and uniformly distributed on $[-1, 1]$. The dependent variable Y , which is the outcome of an experiment with the technical system is defined by

$$Y = m^*(X) + \sigma^* \cdot \lambda^* \cdot \epsilon$$

for some $m^*: \mathbb{R}^d \rightarrow \mathbb{R}$, a noise factor $\sigma^* \in \{0.05, 0.2\}$ and $\lambda^* > 0$ selected as the empirical interquartile range of $m^*(X)$. We define the computer program which simulates the technical system by a functional shift of m^* , i.e.

$$m(x) = m^*(x) + \sigma_m \cdot \lambda_m^{-1} \cdot v_m(x) \quad (x \in \mathbb{R}^d),$$

with $\sigma_m \in \{0.1, 0.2, 0.5\}$, a function $v_m: \mathbb{R}^d \rightarrow \mathbb{R}$ and

$$\lambda_m = \max_{x \in [0, 1]^d} |v_m(x)|.$$

We consider three different models. In each model we use sample sizes $n = 10$, $L_n = 200$, $N_{1,n} = 200$ and $N_{2,n} = 10^4$. The different functions used as m^* are listed below.

$$\begin{aligned} m_1^*(x) &= \cot \left(\frac{\pi}{1 + \exp(x_1^2 + 2 \cdot x_2 + \sin(6 \cdot x_4^3) - 3)} \right) \\ &\quad + \exp(3 \cdot x_3 + 2 \cdot x_4 - 5 \cdot x_5 + \sqrt{x_6 + 0.9 \cdot x_7 + 0.1}) \quad (x \in [0, 1]^7) \\ m_2^*(x) &= \frac{2}{x_1 + 0.008} + 3 \cdot \log(x_2^7 \cdot x_3 + 0.1) \cdot x_4 \quad (x \in [0, 1]^7) \\ m_3^*(x) &= 2 \cdot \log(x_1 \cdot x_2 + 4 \cdot x_3 + |\tan(x_4)| + 0.1) + x_3^4 \cdot x_5^2 \cdot x_6 \\ &\quad - x_4 \cdot x_7 + (3 \cdot x_8^2 + x_9 + 2)^{0.1+4 \cdot x_{10}^2} \quad (x \in [0, 1]^{10}) \end{aligned}$$

As mentioned before, the parameter λ^* is chosen as the empirical interquartile range of $m^*(X)$ calculated on 10^5 realizations of X . The used values are $\lambda_1^* = 9.11$, $\lambda_2^* = 5.68$, $\lambda_3^* = 13.97$ and $\lambda_4^* = 1.64$.

We set

$$\begin{aligned} v_m(x) &= x_1 + \tan(x_2) + x_3^3 + \log(x_4 + 0.1) + 3 \cdot x_5 + x_6 \\ &\quad + \sqrt{x_7 + 0.1} + 0 \cdot \sum_{i=8}^d x_i \quad (x \in [0, 1]^d), \end{aligned}$$

thus $\lambda_m = 8.70$

The density of Y is the convolution of the density of $m(X)$ and a uniform density. We do not try to compute its exact form, instead we compute it approximately by a kernel density estimate (as

implemented in the *MATLAB* routine *ksdensity()* applied to a sample of size 10^6 . In order to evaluate the performance of our density estimates the result is treated as if it were the real density.

The estimates are compared by their L_1 error. Therefore we approximate the integral by a Riemann sum defined on an equidistant partition consisting of 10^4 subintervals. Since we need to take the randomness of the L_1 error into account, we repeat each simulation 50 times and report in Table 2 and Table 3 the median (and in brackets the interquartile range) of the 50 L_1 errors.

Table 2: Median (and interquartile range) of the L_1 error of the four different estimates for the three different models with a constant error in the computer model and five percent noise

σ^* σ_m		5%		
		0.1	0.2	0.5
m_1^*	est. 1	0.704 (0.168)	0.704 (0.168)	0.704 (0.168)
	est. 2	0.702 (0.056)	0.999 (0.0331)	1.344 (0.018)
	est. 3	0.998 (0.345)	0.998 (0.345)	0.998 (0.345)
	est. 4	0.227 (0.167)	0.323 (0.113)	0.570 (0.086)
m_2^*	est. 1	0.525 (0.183)	0.525 (0.183)	0.525 (0.183)
	est. 2	0.702 (0.056)	0.996 (0.038)	1.341 (0.017)
	est. 3	1.086 (0.459)	1.086 (0.459)	1.086 (0.459)
	est. 4	0.216 (0.143)	0.334 (0.119)	0.521 (0.084)
m_3^*	est. 1	0.786 (0.163)	0.786 (0.163)	0.786 (0.163)
	est. 2	0.616 (0.460)	0.935 (0.124)	1.233 (0.263)
	est. 3	1.472 (0.847)	1.472 (0.847)	1.472 (0.847)
	est. 4	0.562 (0.606)	0.835 (0.595)	0.999 (0.499)

Table 3: Median (and interquartile range) of the L_1 error of the three different estimates for the four different models with a constant error in the computer model and twenty percent noise

σ^* σ_m		20%		
		0.1	0.2	0.5
m_1^*	est. 1	0.697 (0.241)	0.697 (0.241)	0.697 (0.241)
	est. 2	0.667 (0.067)	0.470 (0.098)	1.342 (0.019)
	est. 3	1.185 (0.604)	1.185 (0.604)	1.185 (0.604)
	est. 4	0.181 (0.201)	0.304 (0.125)	0.539 (0.096)
m_2^*	est. 1	0.547 (0.181)	0.547 (0.181)	0.547 (0.181)
	est. 2	0.686 (0.061)	0.989 (0.035)	1.340 (0.017)
	est. 3	1.140 (0.401)	1.140 (0.401)	1.140 (0.401)
	est. 4	0.248 (0.233)	0.327 (0.106)	0.511 (0.089)
m_3^*	est. 1	0.666 (0.217)	0.666 (0.217)	0.666 (0.217)
	est. 2	0.579 (0.480)	0.844 (0.229)	1.212 (0.252)
	est. 3	1.263 (0.832)	1.263 (0.832)	1.263 (0.832)
	est. 4	0.573 (0.543)	0.776 (0.499)	0.999 (0.499)

Our newly proposed estimate outperforms the other three estimates in 14 of 18 cases. In all cases if σ_m is sufficiently small our estimate yields a smaller L_1 error than estimates 1 and 3, where the biggest difference is in model four where it is eight times smaller. In any simulation except one it is able to reduce the L_1 error compared to the surrogate estimate on computer model data (est. 2). Due to the complexity of the used model functions m^* and the small sample size of 10 the estimate 3 yields in any simulation the worst results.

We apply the four different estimates on the lateral vibration attenuation system data and illustrate the results in Figure 4. The number of experimental data is equal to 10. To improve the stability of our estimate we increase the sample sizes L_n and $N_{1,n}$ to 500.

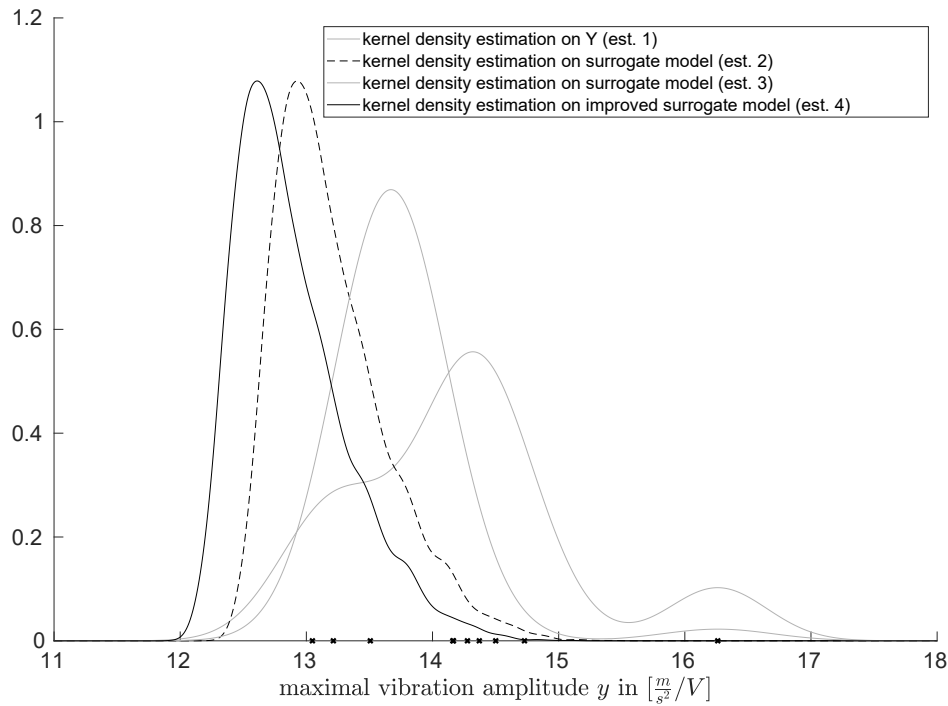


Fig. 4: Four different density estimates and as reference the data set (1) indicated on the x axis.

Obviously the results of our newly proposed estimate are not satisfying. If one observes the positions of the Y values the estimated density of estimate 4 should be shifted to the right in comparison to the estimated density of estimate 2. We guess that this is due to the fact that the input variables are dependent. In the sequel we assume that X is multivariate normally distributed. We estimate its distribution with the *mlest()* routine of the *mvnmle* package of the statistic software *R* and obtain the expectation vector

$$\hat{\mu} = [124.9572 \quad 125.8931 \quad 33046576 \quad 32834749 \quad 0.00678]$$

and the covariance matrix

$$\hat{\Sigma} = \begin{bmatrix} 88.85741 & 32.74759 & 1595777 & -5647359 & 0.0001846703 \\ 32.74759 & 79.76893 & -2919445 & -6593387 & 0.0001762972 \\ 1595777 & -2919445 & 1.070764 \times 10^{12} & 884544431242 & -14.19626 \\ -5647359 & -6593387 & 8.845444 \times 10^{11} & 1.5991 \times 10^{12} & -32.52903 \\ 0.0001846703 & 0.0001762972 & -14.19626 & -32.52903 & 1.600001 \times 10^{-9} \end{bmatrix}.$$

We use the expectation vector and the covariance matrix to generate a data set of size 10^4 of X values. These values are used to evaluate the surrogate models of the estimates 2 to 4. We illustrate the results in Figure 5.

These new results for the estimates 2 and 4 are much more plausible, if one considers the position of the Y values. The result for estimate 3 are not included in the figure, because they were shifted to the negative. But since we only have 10 real data, it is unclear how reliable the estimates 1 and 3 are.

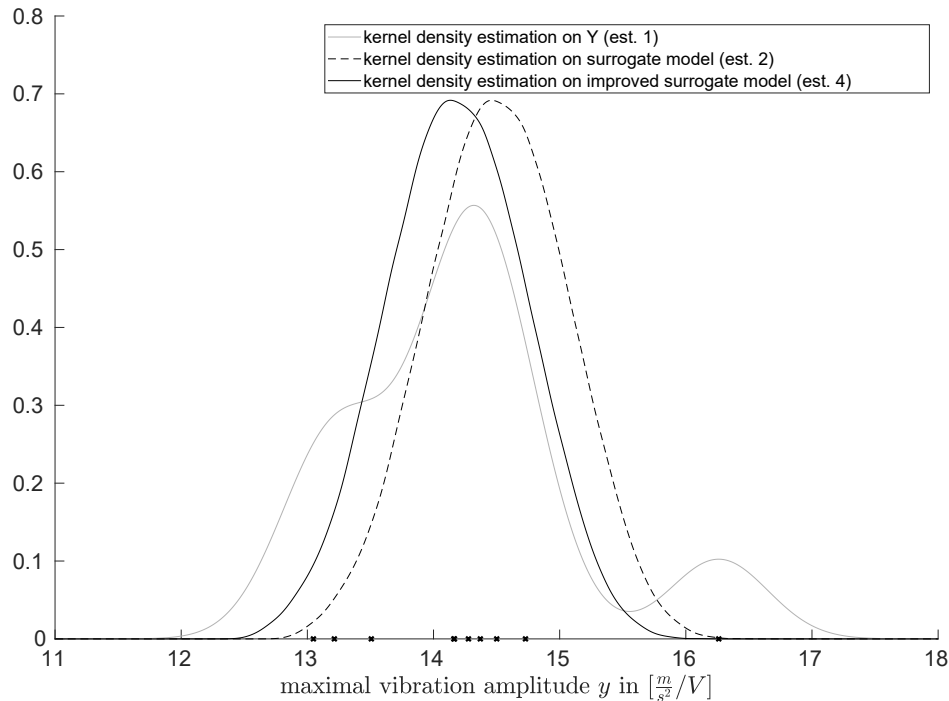


Fig. 5: Four different density estimates, whereas estimates 2 and 4 are evaluated with multivariate estimated input data and as reference the data set (1) indicated on the x axis.

Acknowledgment

The authors would like to thank the German Research Foundation (DFG) for funding this project within the Collaborative Research Center 805.

References

- [1] Coombs, C. H. (1964). *A Theory of Data*. New York, John Wiley & Sons.
- [2] Choi, S.-K., Grandhi, R. V. and Canfield, R. A. (2007). *Reliability-based Structural Design*. Springer-Verlag London Limited.
- [3] Götz, B., Schaeffner, M., Platz, R. and Melz, T. (2016). Lateral vibration attenuation of a beam with circular cross-section by a support with integrated piezoelectric transducers shunted to negative capacitances. *Smart Materials and Structures*, **25.9**, pp. 1–10.
- [4] Kohler, M., and Krzyżak, A. (2017). Nonparametric regression based on hierarchical interaction models. *IEEE Transaction on Information Theory*, **63**, pp. 1620-1630.
- [5] Levenberg, K. (1944). A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Quarterly of Applied Mathematics*, **2**, pp. 164-168.

-
- [6] Marquardt, D. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics*, **11.2**, pp. 431-441.
- [7] Parzen, E. (1962). On the estimation of a probability density function and the mode. *Annals of Mathematical Statistics*, **33**, pp. 1065–1076.
- [8] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, **27**, pp. 832–837.