Estimation of a density from an imperfect simulation model *

Michael Kohler[1] and Adam Krzyżak[2,†]

[1] *Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany, email: kohler@mathematik.tu-darmstadt.de*

[2] *Department of Computer Science and Software Engineering, Concordia University, 1455 De Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8, email: krzyzak@cs.concordia.ca*

February 10, 2017

Abstract

Uncertainty quantification of a technical system can be done using density estimation. The starting point there is usually a stochastic model, which is fitted to the technically system, and the density estimation is done using data from this stochastic model. However, in any application such a stochastic model will not be perfect, and estimation of the density should take into account the inadequacy of the stochastic model. In this paper we show how observed data of the real system together with an imperfect simulation model can be used to derive confidence bands for the density of the technical system. Our main result is that the newly introduced confidence bands allow to derive lower and upper bounds on the probability of intervals in the technical system. Furthermore, we present an upper bound on the area of the confidence band in case of a smooth density. The results are illustrated by applying the estimates to simulated and real data.

AMS classification: Primary 62G07; secondary 62P30.

Key words and phrases: Density estimation, confidence bands, imperfect models, uncertainty quantification.

*Running title: *Estimation from imperfect models*

†Corresponding author. Tel: +1-514-848-2424 ext. 3007, Fax: +1-514-848-2830

1 Introduction

Whenever complex technical systems are designed by engineers, uncertainty has to be taken into account. This uncertainty occurs, e.g., because of the use of an imperfect mathematical model of the technical system during the design process, or because of lack of knowledge about future use. A good quantification of the uncertainty of the system is essential in order to avoid oversizing and to conserve resources.

Uncertainty can be characterized by estimation of quantiles (which enables the characterization of the maximal occurring values) or by estimation of densities (which characterize the occurring randomness completely). In this article the focus will be on density estimation.

The starting point in uncertainty quantification is usually a stochastic model of the technical system. This stochastic model often has parameters which are chosen randomly because their exact values are uncertain and consequently not known, and it computes the outcome of the technical system by computing the value of a function depending on concrete values of the parameters. In case that the distribution of the parameters is known (which we will assume from now on) and that the function, which has to be computed, is given, Monte Carlo can be used to estimate either quantiles or the density of the output of the technical system.

Usually, the stochastic model is evaluated using a computer program, and computer experiments can be used to generate values for the Monte Carlo estimates. However, it often happens that generation of the values is rather time consuming, so that standard Monte Carlo estimates cannot be applied. Instead, one has to apply techniques which are able to quantify the uncertainty in the computer experiment using only a few evaluations of the computer program. There is a vast literature on the design and analysis of such computer experiments, cf., e.g., Santner, Williams, and Notz (2003) and the literature cited therein. Often, so–called surrogate models of the computer experiment are used in order to analyze it. Surrogate models have been introduced and investigated with the aid of the simulated and real data in connection with the quadratic response surfaces in Bucher and Burgund (1990), Kim and Na (1997) and Das and Zheng (2000), in context of support vector machines in Hurtado (2004), Deheeger and Lemaire (2010) and Bourinet, Deheeger and Lemaire (2011), in connection with neural networks in Papadrakakis and Lagaros (2002), and in context of kriging in Kaymaz (2005) and Bichon et al. (2008). Consistency and rate of convergence of density estimates based on surrogate models have been studied in Devroye, Felber and Kohler (2013), Bott, Felber and Kohler (2015) and Felber, Kohler and Krzyżak (2015a). A method for the adaptive choice of the smoothing parameter of such estimates has been presented in Felber, Kohler and Krzyżak (2015b).

Of course, in practice a stochastic model will never be able to represent the real technical system perfectly. So it is clear that the mathematical model is imperfect, and the question is what the consequences of this are in view of uncertainty quantification. The standard approach in science is to make some assumptions about the reality, and to try to quantify the uncertainty under these assumptions. E.g., in Bayesian analysis of computer experiments, Kennedy and O'Hagan (2001), Bayarri et al. (2007), Goh et al. (2013), Han, Santner and Rawlinson (2009), Hidgon et al. (2013) and Wang, Chen and Tsui (2009) model the discrepancy between the computer experiments and the outcome of the technical system by a Gaussian process. Under the assumption that the reality is described by this Gaussian process perfectly, the above papers use this assumption in order to derive confidence intervals of quantiles to be estimated. Here the user can choose a level of the confidence interval, which specifies the probability that the true value is contained in this interval. But of course the latter one is true only if the assumptions about the reality are true, which illustrates the saying "We buy information with assumptions" (Coombs (1964)).

Tuo and Wu (2016) pointed out that the above approach might fail in case of an imperfect computer model, for which there exist no values of the parameters which fit the technical system perfectly, and suggested and analyzed non-Bayesian methods for the choice of the parameters of such models.

Due to the fact that an error on uncertainty quantification of a technical system might result in a fatal error of the real technical system during its usage, it is very important to avoid assumptions in uncertainty quantification as much as possible. For the problem of quantile estimation, Kohler et al. (2017) considered a novel approach towards error estimation in uncertainty quantification, which uses only a very few and rather general assumptions. In particular it was not assumed that the discrepancy between the computer experiments and the outcome of the technical system is a realization of the Gaussian process, and any parametric or nonparametric assumptions on the optimal model describing the technical system have been avoided. Instead observed values of the technical system have been used in order to derive confidence intervals on quantiles of the outcome of the technical system.

In this paper we develop similar methods for the problem of density estimation from an imperfect model. Here the main aim is to construct confidence bands of the density. There are quite a few techniques available in the literature which enable to construct such confidence bands of a density in case that an independent sample of values of the technical system is available, see, e.g., Bickel and Rosenblatt (1973), Giné and Nickl (2010), Hall (1992), Hall and Horowitz (2013), Hall and Titterington (1988), and the literature cited therein. The purpose of this article is to develop a confidence band of a density which improves its quality by using at the same time data from the technical system and data from an imperfect simulation model. In particular we are interested in applications where only a very few data points from the technical system ara available (and where the sample size is so small that nonparametric estimation of a confidence band based alone on these very few data points will not produce good results).

The mathematical model which we consider is as follows: Let (X, Y), (X_1, Y_1), (X_2, Y_2), ... be independent and identically distributed random variables with values in $\mathbb{R}^d \times \mathbb{R}$, and let $m : \mathbb{R}^d \to \mathbb{R}$ be a measurable function. Here Y describes the outcome of an experiment with our technical system, and our aim is to predict the density g of Y (w.r.t. the Lebesgue measure), which we assume to exist. Given

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

(where $L_n, N_n \in \mathbb{N}$) and the data

$$\mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}, \tag{1}$$

our main goal is to construct confidence bands for g. More precisely we want to construct lower and upper bounds $\hat{g}_n^{(lower)}$ and $\hat{g}_n^{(upper)}$ on the density g. Ideally we would like to have

$$\hat{g}_n^{(lower)}(x) \leq g(x) \leq \hat{g}_n^{(upper)}(x) \quad \text{for Lebesgue almost all } x \in \mathbb{R},$$

or equivalently

$$\int_B \hat{g}_n^{(lower)}(x)\, dx \leq \int_B g(x) dx \leq \int_B \hat{g}_n^{(upper)}(x)\, dx \tag{2}$$

for all measurable sets $B \subseteq \mathbb{R}$. The last condition stresses that we are only interested in the density g because it can be used to determine all kinds of probabilities.

In this article we are interested in situations, where the sample size n is rather small (in our application in Section 4 we will have $n = 20$), since the collection of the real data (1) is rather expensive. In view of this it seems hard to construct estimates satisfying (2) for all measurable sets $B \subseteq \mathbb{R}$. Here our model of the technical system is not really helpful, since it is imperfect. So instead, we simplify our goal in the sequel and construct lower and upper bounds $\hat{g}_n^{(lower)}$ and $\hat{g}_n^{(upper)}$ of the density g satisfying

$$\int_I \hat{g}_n^{(lower)}(x)\,dx \le \int_I g(x)dx \le \int_I \hat{g}_n^{(upper)}(x)\,dx \tag{3}$$

for all intervals I with length $|I| \ge \kappa_n$, where $\kappa_n > 0$ is a given value.

In the next section we propose a novel method for the construction of such confidence bands. Our main results are as follows: We show that the newly introduced confidence bands satisfy (3) for finite sample size n and with probability at least $1-\delta$ simultaneously for all intervals of lengths greater than κ_n (where $\delta, \kappa_n > 0$ are given). Furthermore we analyze how the area of the confidence band depends on the smoothness of the density and the error of the surrogate model. The finite sample size behaviour of our estimates is illustrated using simulated data. The usefulness of our newly proposed estimates for uncertainty quantification is demonstrated by using it to analyze the uncertainty occurring in experiments with a suspension strut.

Throughout this paper we use the following notation: $\mathbb{N}$, $\mathbb{N}_0$ and $\mathbb{R}$ are the sets of positive integers, nonnegative integers and real numbers, respectively. If $I \subseteq \mathbb{R}$ is an interval and $\beta \ge 0$, then we denote the length of I by $|I|$ and we define intervals I^β and I_β (where the interval is either extended or shortened on both sides by an interval of length β) by

$$I^\beta = \{x \in \mathbb{R} : [x-\beta, x+\beta] \cap I \ne \emptyset\} \quad \text{and} \quad I_\beta = \{x \in \mathbb{R} : [x-\beta, x+\beta] \subseteq I\}.$$

For $a, b \in \mathbb{R}$ we set

$$[a,b] = \{x \in \mathbb{R} : a \le x \le b\} \quad \text{and} \quad (a,b) = \{x \in \mathbb{R} : a < x < b\},$$

similarly we define $(a,b]$ and $[a,b)$. Let $p = k + \beta$ for some $k \in \mathbb{N}_0$ and $0 < \beta \le 1$, and let $C > 0$. A function $m : \mathbb{R}^d \to \mathbb{R}$ is called (p,C)-smooth, if for every $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$ with $\sum_{j=1}^d \alpha_j = k$ the partial derivative $\frac{\partial^k m}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}$ exists and satisfies

$$\left| \frac{\partial^k m}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}(x) - \frac{\partial^k m}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}(z) \right| \le C \cdot \|x - z\|^\beta$$

for all $x, z \in \mathbb{R}^d$. If X is a random variable, then $\mathbf{P}_X$ is the corresponding distribution, i.e., the measure associated with the random variable.

The outline of this paper is as follows: In Section 2 the construction of the confidence bands is explained. The main results are presented in Section 3 and proven in Section 5. The finite sample size perfomance of our confidence bands is illustrated in Section 4 by applying it to simulated and real data.

2 A new method for the construction of confidence bands for densities

In this section we describe our ideas behind the construction of confidence bands for densities.

In order to construct confidence bands satisfying (3), we proceed as follows: If we assume that we know the N_n outcomes $Y_{n+L_n+1}, \dots, Y_{n+L_n+N_n}$ of the experiments with the technical system corresponding to the (random) parameters $X_{n+L_n+1}, \dots, X_{n+L_n+N_n}$, then we know from empirical process theory that the empirical distribution

$$\hat{\mu}_{Y,N_n}(A) = \frac{1}{N_n} \sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{Y_i \in A\}} \quad (A \subseteq \mathbb{R})$$

satisfies

$$\int_I g(y)\, dy = \mathbf{P}\{Y \in I\} \leq \hat{\mu}_{Y,N_n}(I) + \frac{2 \cdot \sqrt{\log N_n}}{\sqrt{N_n}}$$

and

$$\int_I g(y)\, dy = \mathbf{P}\{Y \in I\} \geq \hat{\mu}_{Y,N_n}(I) - \frac{2 \cdot \sqrt{\log N_n}}{\sqrt{N_n}}$$

uniformly for all intervals $I \subseteq \mathbb{R}$ with a large probability.

Our first idea is to use our simulation model to derive bounds on $\hat{\mu}_{Y,N_n}(I)$. To do this, we use the data

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

in order to construct a surrogate estimate

$$m_n(\cdot) = m_n(\cdot, (X_{n+1}, m(X_{n+1})), \dots, (X_{n+L_n}, m(X_{n+L_n}))) : \mathbb{R}^d \to \mathbb{R}$$

of m. Then we compute

$$m_n(X_{n+L_n+1}), \dots, m_n(X_{n+L_n+N_n}) \tag{4}$$

and define the corresponding empirical distribution $\hat{\mu}_{m_n(X),N_n}$ by

$$\hat{\mu}_{m_n(X),N_n}(A) = \frac{1}{N_n} \sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{m_n(X_i) \in A\}} \quad (A \subseteq \mathbb{R}).$$

We use this empirical distribution in order to derive upper and lower bounds on $\hat{\mu}_{Y,N_n}(I)$. The starting point here is the observation that if we define for $\beta \geq 0$ and $I \subseteq \mathbb{R}$

$$I^\beta = \{x \in \mathbb{R} : [x-\beta, x+\beta] \cap I \neq \emptyset\} \quad \text{and} \quad I_\beta = \{x \in \mathbb{R} : [x-\beta, x+\beta] \subseteq I\},$$

then we have

$$\hat{\mu}_{Y,N_n}(I) \le \hat{\mu}_{m_n(X),N_n}(I^\beta) + \frac{1}{N_n} \sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i - m_n(X_i)| > \beta\}}$$

and

$$\hat{\mu}_{Y,N_n}(I) \ge \hat{\mu}_{m_n(X),N_n}(I_\beta) - \frac{1}{N_n} \sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i - m_n(X_i)| > \beta\}}$$

(see proof of Lemma 3 below). Hence as soon we know a bound on

$$\frac{1}{N_n} \sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i - m_n(X_i)| > \beta\}}$$

for a suitable value of β, we can use $\hat{\mu}_{m_n(X),N_n}$ to derive upper and lower bounds on $\hat{\mu}_{Y,N_n}$.

In order to derive such a bound we use (X_1, Y_1), ..., (X_n, Y_n). Here we apply a result from Kohler et al. (2017), which tells us that if we set

$$\hat{\beta} = \max_{i=1,\dots,n} |Y_i - m_n(X_i)|$$

and choose some $\epsilon_n > 0$, then

$$\mathbf{P}\left\{ |Y - m_n(X)| > \hat{\beta} \middle| X_1, \dots, X_{n+L_n}, Y_1, \dots, Y_n \right\} \le \epsilon_n$$

holds outside of an event whose probability is bounded from above by $(1-\epsilon_n)^n$ (cf., Lemma 2 below). Together with the inequality of Hoeffding this enables us to show that for suitable chosen $\epsilon_n, \gamma_n > 0$

$$\frac{1}{N_n} \sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i - m_n(X_i)| > \hat{\beta}\}} \le \epsilon_n + \gamma_n$$

holds with large probability (cf., Lemma 1 below).

Finally we use the kernel density estimate

$$\hat{f}_{m_n(X),N_n,h_{N_n}}(y) = \frac{1}{N_n \cdot h_{N_n}} \sum_{i=n+L_n+1}^{n+L_n+N_n} K\left(\frac{y - m_n(X_i)}{h_{N_n}} \right)$$

corresponding to data (4) in order to derive upper and lower bounds on $\hat{\mu}_{m_n(X),N_n}$. Since we have are given data (4), we can compute the difference between probabilities computed via $\hat{f}_{m_n(X),N_n,h_{N_n}}$ and $\hat{\mu}_{m_n(X),N_n}$ and include these differences into our upper and lower bounds.

This way with large probability for any interval $I \subseteq \mathbb{R}$ the following upper and lower bounds are valid:

$$\begin{aligned}\int_I g(y)\,dy &\leq \int_I \hat{f}_{m_n(X),N_n,h_{N_n}}(y)\,dy + \epsilon_n + \gamma_n + \frac{2\cdot\sqrt{\log N_n}}{\sqrt{N_n}}\\ &\quad + \left(\hat{\mu}_{m_n(X),N_n}(I^{\hat{\beta}}) - \int_I \hat{f}_{m_n(X),N_n,h_{N_n}}(y)\,dy\right)\end{aligned}$$

and

$$\begin{aligned}\int_I g(y)\,dy &\geq \int_I \hat{f}_{m_n(X),N_n,h_{N_n}}(y)\,dy - \epsilon_n - \gamma_n - \frac{2\cdot\sqrt{\log N_n}}{\sqrt{N_n}}\\ &\quad - \left(\int_I \hat{f}_{m_n(X),N_n,h_{N_n}}(y)\,dy - \hat{\mu}_{m_n(X),N_n}(I_{\hat{\beta}})\right).\end{aligned}$$

Finally we use these bounds to derive estimates satisfying (3). In view of the bounds above, it suffices to choose $\hat{g}_n^{(lower)}$ and $\hat{g}_n^{(upper)}$ such that we have for any interval $I \subseteq \mathbb{R}$ with length $|I| \geq \kappa_n$

$$\begin{aligned}\int_I \hat{g}_n^{(upper)}(y)\,dy &\geq \int_I \hat{f}_{m_n(X),N_n,h_{N_n}}(y)\,dy + \epsilon_n + \gamma_n + \frac{2\cdot\sqrt{\log N_n}}{\sqrt{N_n}}\\ &\quad + \left(\hat{\mu}_{m_n(X),N_n}(I^{\hat{\beta}}) - \int_I \hat{f}_{m_n(X),N_n,h_{N_n}}(y)\,dy\right)\end{aligned}$$

and

$$\begin{aligned}\int_I \hat{g}_n^{(lower)}(y)\,dy &\leq \int_I \hat{f}_{m_n(X),N_n,h_{N_n}}(y)\,dy - \epsilon_n - \gamma_n - \frac{2\cdot\sqrt{\log N_n}}{\sqrt{N_n}}\\ &\quad - \left(\int_I \hat{f}_{m_n(X),N_n,h_{N_n}}(y)\,dy - \hat{\mu}_{m_n(X),N_n}(I_{\hat{\beta}})\right).\end{aligned}$$

To achieve this, we define

$$\begin{aligned}\hat{g}_n^{(upper)}(y) &= \hat{f}_{m_n(X),N_n,h_{N_n}}(y) + \frac{\epsilon_n + \gamma_n + \frac{2\cdot\sqrt{\log N_n}}{\sqrt{N_n}}}{\kappa_n}\\ &\quad + \frac{1}{\kappa_n}\cdot \sup_{\substack{J\, interval,\, y\in J,\\ |J|>\kappa_n}} \left(\hat{\mu}_{m_n(X),N_n}(J^{\hat{\beta}}) - \int_J \hat{f}_{m_n(X),N_n,h_{N_n}}(t)\,dt\right)\end{aligned}$$

and

$$\begin{aligned}\hat{g}_n^{(lower)}(y) &= \hat{f}_{m_n(X),N_n,h_{N_n}}(y) - \frac{\epsilon_n + \gamma_n + \frac{2\cdot\sqrt{\log N_n}}{\sqrt{N_n}}}{\kappa_n}\\ &\quad - \frac{1}{\kappa_n}\cdot \sup_{\substack{J\, interval,\, y\in J,\\ |J|>\kappa_n}} \left(\int_J \hat{f}_{m_n(X),N_n,h_{N_n}}(t)\,dt - \hat{\mu}_{m_n(X),N_n}(J_{\hat{\beta}})\right).\end{aligned}$$

3 Main results

Theorem 1 *Let* (X,Y)*,* (X_1,Y_1)*, ... be* $\mathbb{R}^d \times \mathbb{R}$*–valued random variables and let* $m : \mathbb{R}^d \to \mathbb{R}$ *be a measurable function. Assume that* Y *has a density* g *with respect to the Lebesgue-Borel-measure. Let* $n, N_n, L_n \in \mathbb{N}$*,* $h_{N_n}, \kappa_n > 0$ *and* $\delta, \epsilon_n \in (0,1)$*, and let*

$$m_n(\cdot) = m_n(\cdot, (X_{n+1}, m(X_{n+1})), \dots, (X_{n+L_n}, m(X_{n+L_n}))) : \mathbb{R}^d \to \mathbb{R}.$$

Assume $(1-\epsilon_n)^n + 2/N_n^2 < \delta$*. Set*

$$\hat{\beta} = \max_{i=1,\dots,n} |Y_i - m_n(X_i)|,$$

and

$$\gamma_n = \sqrt{\frac{-\ln\left(\delta - 2/N_n^2 - (1-\epsilon_n)^n\right)}{2 \cdot N_n}}.$$

Define

$$\hat{f}_{m_n(X),N_n,h_{N_n}}(y) = \frac{1}{N_n \cdot h_{N_n}} \sum_{i=n+L_n+1}^{n+L_n+N_n} K\left(\frac{y - m_n(X_i)}{h_{N_n}}\right),$$

$$\hat{\mu}_{m_n(X),N_n}(A) = \frac{1}{N_n} \sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{m_n(X_i) \in A\}} \quad (A \subseteq \mathbb{R}),$$

$$\begin{aligned} \hat{g}_n^{(upper)}(y) &= \hat{f}_{m_n(X),N_n,h_{N_n}}(y) + \frac{1}{\kappa_n} \cdot \left(\epsilon_n + \gamma_n + \frac{2 \cdot \sqrt{\log N_n}}{\sqrt{N_n}}\right) \\ &\quad + \sup_{\substack{J\, interval,\, y \in J \\ |J| > \kappa_n}} \frac{\hat{\mu}_{m_n(X),N_n}(J^{\hat{\beta}}) - \int_J \hat{f}_{m_n(X),N_n,h_{N_n}}(t)\, dt}{\kappa_n} \end{aligned}$$

and

$$\begin{aligned} \hat{g}_n^{(lower)}(y) &= \hat{f}_{m_n(X),N_n,h_{N_n}}(y) - \frac{1}{\kappa_n} \cdot \left(\epsilon_n + \gamma_n + \frac{2 \cdot \sqrt{\log N_n}}{\sqrt{N_n}}\right) \\ &\quad - \sup_{\substack{J\, interval,\, y \in J \\ |J| > \kappa_n}} \frac{\int_J \hat{f}_{m_n(X),N_n,h_{N_n}}(t)\, dt - \hat{\mu}_{m_n(X),N_n}(J_{\hat{\beta}})}{\kappa_n} \end{aligned}$$

Then with probability at least $1-\delta$*, the following inequality holds simultaneously for all intervals* I *with length* $|I| > \kappa_n$*:*

$$\int_I \hat{g}_{N_n}^{(lower)}(x)\, dx \le \int_I g(y)\, dy \le \int_I \hat{g}_{N_n}^{(upper)}(x)\, dx. \quad (5)$$

Remark 1. It follows from the proof of Theorem 1 that (5) holds with probability at least $1-\delta$ simultaneously for all intervals I with length $|I| > \kappa_n$ and for all bandwidths $h_{N_n} > 0$.

Remark 2. If g is (p, C)–smooth for some $p \in (0, 1]$ we have for any interval $I \subseteq \mathbb{R}$ and any $y \in I$

$$g(y) - C \cdot |I|^p \leq \frac{1}{|I|} \cdot \int_I g(t)\, dt \leq g(y) + C \cdot |I|^p$$

which implies

$$-C \cdot |I|^p + \frac{1}{|I|} \cdot \int_I g(t)\, dt \leq g(y) \leq C \cdot |I|^p + \frac{1}{|I|} \cdot \int_I g(t)\, dt.$$

Consequently in this case we can conclude from (5) that we have for all $y \in \mathbb{R}$ and all intervals I satisfying $y \in I$ and $|I| > \kappa_n$

$$-C \cdot |I|^p + \frac{1}{|I|} \cdot \int_I \hat{g}_n^{(lower)}(t)\, dt \leq g(y) \leq C \cdot |I|^p + \frac{1}{|I|} \cdot \int_I \hat{g}_n^{(upper)}(t)\, dt.$$

From this we can conclude for any $y \in \mathbb{R}$:

$$\begin{aligned} &-C \cdot \kappa_n^p + \frac{1}{\kappa_n} \cdot \max\left\{ \int_{y-\kappa_n}^{y} \hat{g}_n^{(lower)}(t)\, dt, \int_{y}^{y+\kappa_n} \hat{g}_n^{(lower)}(t)\, dt \right\} \\ &\leq g(y) \leq C \cdot \kappa_n^p + \frac{1}{\kappa_n} \cdot \min\left\{ \int_{y-\kappa_n}^{y} \hat{g}_n^{(upper)}(t)\, dt, \int_{y}^{y+\kappa_n} \hat{g}_n^{(upper)}(t)\, dt \right\}. \end{aligned}$$

Hence in case of a (p, C)–smooth density we can derive from (5) also pointwise bounds on the density provided we know the smoothness of the density, i.e., the values of (p, C)).

It is an open problem whether one can derive from (5) also similar bounds in case that (p, C) is unknown.

In our next result we analyze the area between the upper and the lower bound on the density given in Theorem 1.

Theorem 2 *Assume that the assumptions of Theorem 1 hold and that, in addition, the density g of Y is (p, C)–smooth for some $C > 0$, $p \in (0, 1]$ and that the support of g is compact. Let $\delta > 0$ be arbitrary, set*

$$\epsilon_n = \frac{\log n}{n}$$

and

$$\gamma_n = \sqrt{\frac{-\log\left(\delta - \frac{2}{N_n^2} - (1-\epsilon_n)^n\right)}{2 \cdot N_n}},$$

and choose N_n and h_{N_n} such that

$$N_n \to \infty \quad (n \to \infty) \quad \textit{and} \quad \limsup_{n\to\infty} h_{N_n} < \infty.$$

Let $K : \mathbb{R} \to \mathbb{R}$ be a symmetric and bounded density with compact support, which is monotonically decreasing on $\mathbb{R}_+$, and define $\hat{g}_n^{(upper)}$ and $\hat{g}_n^{(lower)}$ as in Theorem 1.

Then there exists a constant $c_1 > 0$ such that outside of an event, whose probability tends to δ for $n \to \infty$, we have for any interval $I \subseteq \mathbb{R}$ with length $|I| < \infty$:

$$\begin{aligned}
&\int_I \left|\hat{g}_n^{(upper)}(y) - \hat{g}_n^{(lower)}(y)\right| dy \qquad (6)\\
&\leq \frac{|I|}{\kappa_n} \cdot \left(c_1 \cdot \left(h_{N_n}^p + \frac{1}{\sqrt{N_n \cdot h_{N_n}}}\right) + 8 \cdot \frac{\log n}{n} + 18 \cdot \frac{\sqrt{\log N_n}}{\sqrt{N_n}}\right.\\
&\qquad \left. + \frac{8 \cdot (K(0) + \|g\|_\infty) \cdot \max_{i=1,\dots,n} |Y_i - m_n(X_i)|}{\min\{h_{N_n}, 1\}}\right).
\end{aligned}$$

In particular, in case that we have

$$\max_{i=1,\dots,n} |Y_i - m_n(X_i)| \to 0 \quad (n \to \infty), \qquad (7)$$

we can set

$$h_{N_n} = c_2 \cdot \left(\max_{i=1,\dots,n} |Y_i - m_n(X_i)|\right)^{1/(p+1)}$$

and choose N_n sufficiently large, and can conclude from inequality (6) that

$$\begin{aligned}
&\int_I \left|\hat{g}_n^{(upper)}(y) - \hat{g}_n^{(lower)}(y)\right| dy\\
&\leq c_3 \cdot \frac{|I|}{\kappa_n} \cdot \left(\frac{\log n}{n} + \left(\max_{i=1,\dots,n} |Y_i - m_n(X_i)|\right)^{p/(p+1)}\right).
\end{aligned}$$

Remark 3. Assumption (7) requires that either $Y = m(X)$ holds or that our approximation $m(X)$ of Y is changed with increasing sample size n and is becoming better and better for $n \to \infty$.

4 Application to simulated and real data

In this section we illustrate the finite sample size performance of our estimates by applying them to simulated and real data.

We start with an application to simulated data, where we illustrate how the size of the error of the model influences the performance of our estimates. To do this, we choose X d-dimensional standard normally distributed and ϵ uniformly distributed on $[0, 1]$ such that X and ϵ are independent, set

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

for some $m : \mathbb{R}^d \to \mathbb{R}$ defined below and $\sigma \in \{0.1, 0.5\}$, and let (X_1, Y_1), (X_2, Y_2), ... be independent and identically distributed random variables. Our estimate gets

$$(X_1, Y_1), \ldots, (X_n, Y_n)$$

as data from the real technical system,

$$(X_{n+1}, m(X_{n+1})), \ldots, (X_{n+L_n}, m(X_{n+L_n}))$$

as data from the (imperfect) model (where σ controls the maximal error occurring in this model), and the additional X-values

$$X_{n+L_n+1}, \ldots, X_{n+L_n+N_n}.$$

In all of our applications we choose $n = 20$, $L_n = 500$ and $N_n = 500,000$. As surrogate estimate we use a thin plate spline as implemented in the routine *Tps*() in the statistics software *R*, with smoothing parameter chosen by generalized cross validation. Our estimate uses the naive kernel

$$K(x) = \frac{1}{2} \cdot I_{\{x \in [-1,1]\}}$$

and is implemented such that it considers in its computation only intervals with endpoints on some grid consisting of 2001 points chosen equidistantly in some interval depending on the function m (see below). For the bandwidth h_{N_n} we choose the value produced by the procedure *density*() in *R* applied to the N_n data points (4) (which is chosen by L_2 cross validation for the kernel density estimate). In case of the simulated data below we choose $\kappa_n = 2$.

The density of Y is the convolution of the density of $m(X)$ and uniform density. We do not try to compute its exact form, instead we compute it approximately by applying a kernel density estimate (as implemented in the routine *density*() in *R*) to a sample of size $1,000,000$ of Y. In order to judge the quality of our confidence band the resulting density is treated in our simulations as if it is the real density.

In our first model we choose $d = 2$ and

$$m(x_1, x_2) = 2 \cdot x_1 + x_2 + 2.$$

The interval on which we compute our estimate is chosen as $[-5, 9]$. Typical

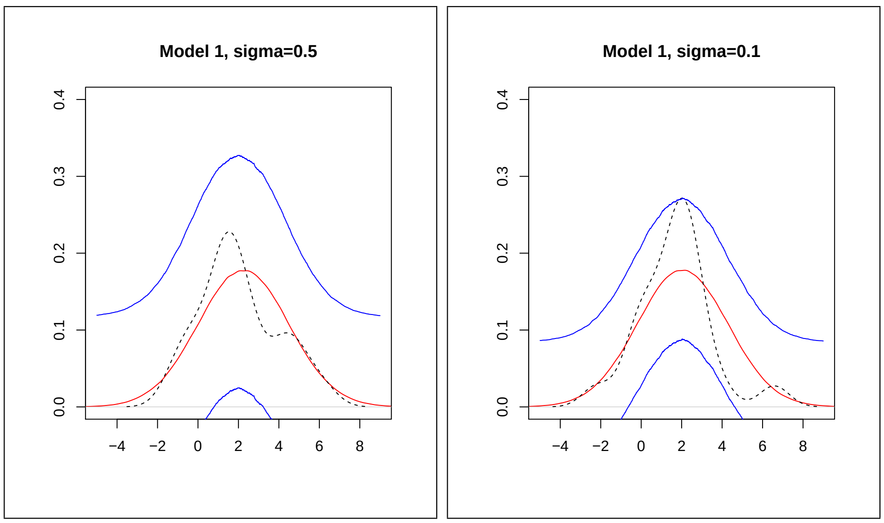

Figure 1: Typical simulations in the first model (with $\sigma = 0.5$ in the left panel and $\sigma = 0.1$ in the right panel). The dashed curve is the kernel density estimate applied only to the $n = 20$ real data points, the black curve a density estimate applied to a sample of size $1,000,000$ of Y and is considered as a very good approximation to the true density g. The confidence band is drawn in blue.

simulations for $\sigma = 0.5$ and $\sigma = 0.1$ are shown in Figure 1. As can be seen from Figure 1, in case of the smaller error of the model (i.e., $\sigma = 0.1$ instead of $\sigma = 0.5$) the confidence band is narrower, and has about the same distance from our reference density as the kernel density estimate applied to the sample of size $n = 20$ of Y. Of course, the data points are random. To confirm the above observation, we repeat the above simulations 50 times and compute the median and the interquartile range of the maximal area between the confidence band over any interval of length $\kappa_n = 2$. We compare this value with the maximal area over any interval of length $\kappa_n = 2$ between the density estimate based on the $n = 20$ data points of Y and our reference density (which we consider as a lower bound on the corresponding value for any confidence band based on a kernel density estimate using $n = 20$ data points of Y). In case $\sigma = 0.5$ we get for our confidence band as median value 0.613 (with IQR 0.020) while the density estimate with $n = 20$ real data points achieves a smaller median value 0.345 (with IQR 0.00067). But in case of the smaller error, i.e., $\sigma = 0.1$, our estimated confidence band narrows and achieves with median 0.368 (and IQR 0.00061) approximately the same value as the density estimate with $n = 20$ real data points.

In our second model we choose again $d = 2$, but this time m is defined by

$$m(x_1, x_2) = x_1^2 + x_2^2,$$

and the interval on which we compute our estimate is $[-1, 10]$. Typical

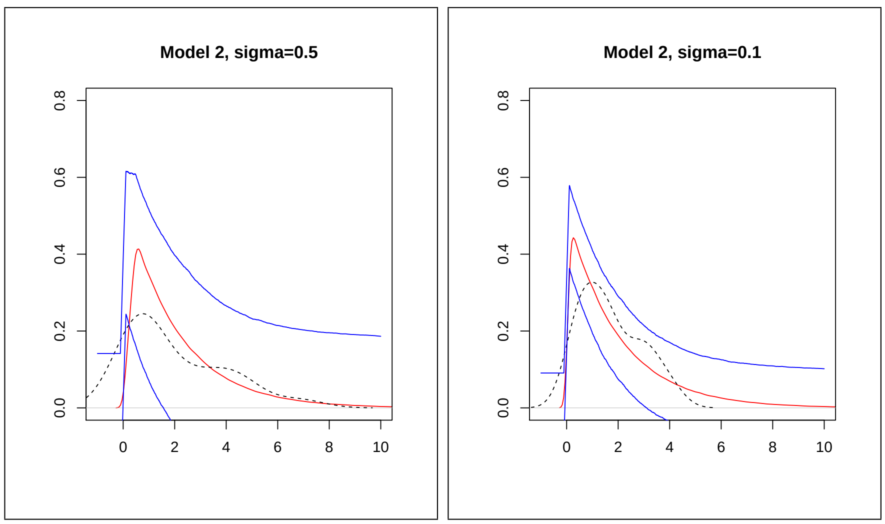

Figure 2: Typical simulations in the second model (with $\sigma = 0.5$ in the left panel and $\sigma = 0.1$ in the right panel). The dashed curve is the kernel density estimate applied only to the $n = 20$ real data points, the black curve a density estimate applied to a sample of size $1,000,000$ of Y and is considered as a very good approximation to the true density g. The confidence band is drawn in blue.

simulations for $\sigma = 0.5$ and $\sigma = 0.1$ are shown in Figure 2. Repeating these simulations again 50 times, we get in case of $\sigma = 0.5$ for the maximal area between our confidence band over any interval of length $\kappa_n = 2$ the median value 0.890 (with IQR 0.205), while the density estimate with $n = 20$ real data points achieves a smaller median value of 0.614 (with IQR 0.00063). But in case of the smaller error, i.e., $\sigma = 0.1$, our estimated confidence band narrows and achieves with median 0.440 (and IQR 0.270) now a smaller value than the density estimate with $n = 20$ real data points.

This effect is in our third model even stronger. Here we choose $d = 1$, define m by

$$m(x) = \exp(x),$$

and compute our estimate again on the interval $[-1, 10]$. Typical simulations

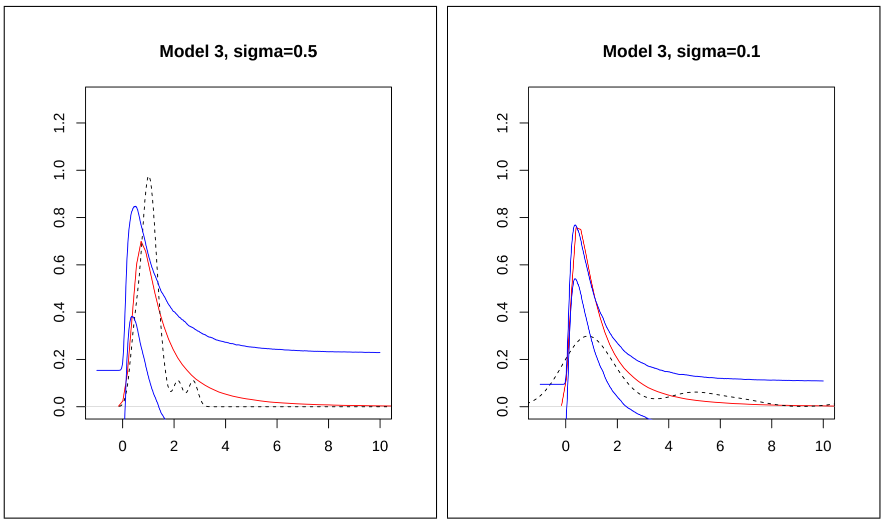

Figure 3: Typical simulations in the third model (with $\sigma = 0.5$ in the left panel and $\sigma = 0.1$ in the right panel). The dashed curve is the kernel density estimate applied only to the $n = 20$ real data points, the black curve a density estimate applied to a sample of size $1,000,000$ of Y and is considered as a very good approximation to the true density g. The confidence band is drawn in blue.

for $\sigma = 0.5$ and $\sigma = 0.1$ are shown in Figure 3. Repeating these simulations again 50 times, we get in case of $\sigma = 0.5$ for the maximal area between our confidence band on any interval of length $\kappa_n = 2$ the median value 1.010 (with IQR 0.047), while the density estimate with $n = 20$ real data points achieves a smaller median value 0.756 (with IQR 0.00061). But in case of the smaller error, i.e., $\sigma = 0.1$, our estimated confidence band narrows and achieves with median 0.455 (and IQR 0.014) now a much smaller value than the density estimate with $n = 20$ real data points.

Summarizing the results of our simulations we see that as indicated by the bound presented in Theorem 2 the area of our newly introduced confidence band is influenced drastically by the size of the error of the model, that this effect is clearly visible for finite sample size, and that for suitably chosen functions m and small enough error of the model the newly proposed confidence band is closer to the true density than the kernel density estimate based only on n real data points. The last point suggests that even an imperfect simulation model is helpful in constructing confidence bands for a density.

Finally we illustrate the usefulness of our newly proposed method for uncertainty quantification by using it to analyze the uncertainty occurring in experiments with a suspension strut (cf., Figure 4), which serves as an academic demonstrator to study uncertainty in load distributions and the ability to control vibrations, stability and load paths in suspension struts such as aircarft landing gears. A CAD illustration of this suspension strut can be

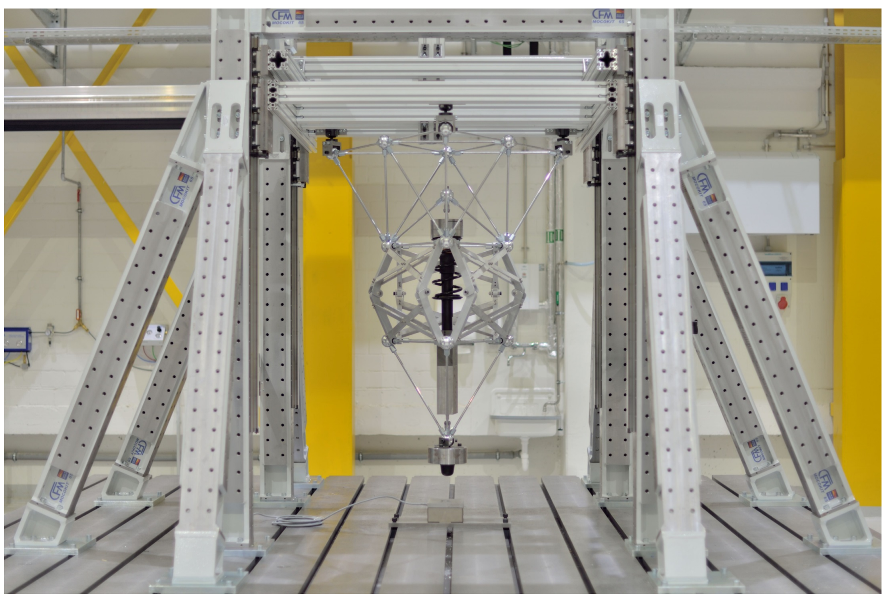

Figure 4: A photo of the demonstrator of a suspension strut and its experimental test setup.

found in Figure 5 (left). This suspension strut consists of an upper and lower structure, where the lower structure contains a spring–damper component. The spring–damper component transmits the axial forces between the upper and lower structures of the suspension strut. The aim of our analysis is the analysis of the behaviour of the maximum relative compression of the spring damper component in case that the free fall height is chosen randomly. Here we assume that the free fall heights are independent normally distributed

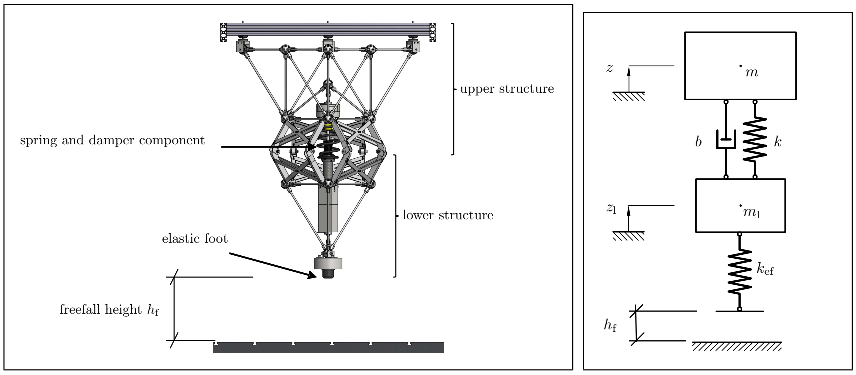

Figure 5: A CAD illustration of the suspension strut (left) and illustration of a simplified model of the suspension strut (right).

with mean 0.05 meter and standard deviation 0.0057 meter.

We use the results of $n = 20$ corresponding experiments. From this data we can estimate the density of the maximum relative compression. In the sequel we improve this density estimate by using, in addition, also data corresponding to a simplified mathematical model of the suspension strut (cf., Figure 5 (right)). Here the upper and the lower structures of the suspension strut are two lump masses m and m_1, the spring damper component is represented by a stiffness parameter k and a suitable damping coefficient b, and the elastic foot is represented by another stiffness parameter k_{ef}. Using a nonlinear stiffness and a velocity dependent damping it is possible to compute the maximum relative compression by solving a differential equation using Runge-Kutta algorithm (cf., Mallapur and Platz (2016)). We use the results of $L_n = 500$ corresponding computer experiments to construct a surrogate estimate m_n as described above. Here, our model is imperfect: if we evaluate m_n on the x–values of our experimental data, and compare the values of m_n with the measured values of the maximum relative compression, we observe that the maximal absolute error is $\hat{\beta} = 0.00067$ and is definitely not equal to zero.

The techniques developed in this paper enable us to quantify the influence of this error on the density estimate. Figure 6 shows the confidence band for the density produced by our method together with an estimate of the density based on $N_n = 500,000$ data points from the surrogate model (red).

For the computation of the confidence band we use $N_n = 500,000$, $h \in \{0.0001, 0.0006, 0.001, 0.005, 0.01\}$, $\delta = 0.05$, $\kappa = 0.005$ and $\epsilon = 0.14$ and compute the minimum and the maximum of the corresponding 5 upper

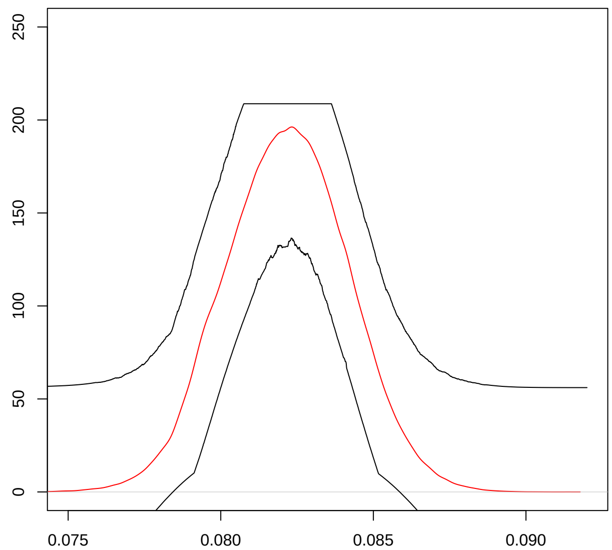

Figure 6: The confidence band for the density together with a density estimate (red) based on $N_n = 500,000$ points from the surrogate model.

and lower bounds of the density, resp. We simplify the computation of the confidence band by considering only intervals which have endpoints on a grid with grid size 0.00001.

From Figure 6 we see that the inadequacy of our model results in uncertainty concerning the true density in our technical system, and we can decide whether this uncertainty is acceptable for our application or whether we should try to improve the model of the technical system in order to reduce this uncertainty.

5 Proofs

5.1 Proof of Theorem 1

Lemma 1 *Let (X, Y), (X_1, Y_1), $\ldots$ be $\mathbb{R}^d \times \mathbb{R}$–valued random variables and let $m : \mathbb{R}^d \to \mathbb{R}$ be a measurable function. Let $n, N_n, L_n \in \mathbb{N}$, $h_{N_n}, \eta_{N_n} > 0$*

and $\delta, \epsilon_n \in (0,1)$, *and let*

$$m_n(\cdot) = m_n(\cdot, (X_{n+1}, m(X_{n+1})), \dots, (X_{n+L_n}, m(X_{n+L_n}))) : \mathbb{R}^d \to \mathbb{R}.$$

Assume $(1-\epsilon_n)^n + 2/N_n^2 < \delta$. *Set*

$$\hat{\beta} = \max_{i=1,\dots,n} |Y_i - m_n(X_i)|,$$

and

$$\gamma_n = \sqrt{\frac{-\ln\left(\delta - 2/N_n^2 - (1-\epsilon_n)^n\right)}{2 \cdot N_n}}.$$

Then

$$\mathbf{P}\left\{ \frac{1}{N_n} \sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i - m_n(X_i)| > \hat{\beta}\}} > \epsilon_n + \gamma_n \right\} \leq \delta - 2/N_n^2.$$

In the proof we will need the following result from Kohler et al. (2017).

Lemma 2 *Let* (X,Z), (X_1,Z_1), $\dots$, (X_n,Z_n) *be independent and identically distributed* $\mathbb{R}^d \times \mathbb{R}$*–valued random variables, let* $K_n \subseteq \mathbb{R}^d$ *and let* $\epsilon_n \in (0,1)$ *be arbitrary. Then*

$$\mathbf{P}_{(X,Z)}\left(\left\{ (x,z) \in K_n \times \mathbb{R} \,:\, z > \max_{\substack{i=1,\dots,n,\\ X_i \in K_n}} Z_i \right\} \right) \leq \epsilon_n \qquad (8)$$

holds outside of an event, whose probability is bounded from above by $(1-\epsilon_n)^n$.

Proof. See Lemma 3 in Kohler et al. (2017). □

Proof of Lemma 1. Set

$$\bar{\mathcal{D}}_n = \{(X_1,Y_1), \dots, (X_n,Y_n), X_{n+1}, \dots, X_{n+L_n}\}.$$

Application of Lemma 2 (with $Z = |Y - m_n(X)|$ and $K_n = \mathbb{R}^d$) yields

$$\mathbf{P}\left\{ \frac{1}{N_n} \sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i - m_n(X_i)| > \hat{\beta}\}} > \epsilon_n + \gamma_n \right\}$$
$$\leq \mathbf{P}\left\{ \mathbf{P}\left\{ |Y - m_n(X)| > \hat{\beta} \middle| \bar{\mathcal{D}}_n \right\} > \epsilon_n \right\}$$

$$+\mathbf{P}\left\{\frac{1}{N_n}\sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i-m_n(X_i)|>\hat{\beta}\}} \right.$$

$$\left. -\mathbf{P}\left\{|Y-m_n(X)|>\hat{\beta}\middle|\bar{\mathcal{D}}_n\right\}>\gamma_n\right\}$$

$$\leq (1-\epsilon_n)^n + \mathbf{P}\left\{\frac{1}{N_n}\sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i-m_n(X_i)|>\hat{\beta}\}} \right.$$

$$\left. -\mathbf{P}\left\{|Y-m_n(X)|>\hat{\beta}\middle|\bar{\mathcal{D}}_n\right\}>\gamma_n\right\}.$$

By the inequality of Hoeffding (cf., e.g., Lemma A.3 in Györfi et al. (2002)) applied conditional on $\bar{\mathcal{D}}_n$ the last term can be bounded from above by

$$(1-\epsilon_n)^n + \exp\left(-2\cdot N_n\cdot\gamma_n^2\right) = \delta - 2/N_n^2,$$

where the last equality follows from the definition of γ_n. □

Lemma 3 *Let*

$$\hat{f}_{m_n(X),N_n,h_{N_n}}(x) = \frac{1}{N_n\cdot h_{N_n}}\sum_{i=n+L_n+1}^{n+L_n+N_n} K\left(\frac{x-m_n(X_i)}{h_{N_n}}\right) \quad (x\in\mathbb{R}),$$

$$\hat{\mu}_{m_n(X),N_n}(A) = \frac{1}{N_n}\sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{m_n(X_i)\in A\}} \quad (A\subseteq\mathbb{R})$$

and

$$\hat{\mu}_{Y,N_n}(A) = \frac{1}{N_n}\sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{Y_i\in A\}} \quad (A\subseteq\mathbb{R}).$$

Let $G_Y(y) = \mathbf{P}\{Y\leq y\}$ $(y\in\mathbb{R})$ be the cdf. of Y and let $\hat{G}_{Y,N_n} = \hat{\mu}_{Y,N_n}((-\infty,y])$ $(y\in\mathbb{R})$ be the empirical cdf. of Y.

Then we have for any interval $I\subseteq\mathbb{R}$ and any $\beta\geq 0$:

$$\mathbf{P}\{Y\in I\} \leq \int_I \hat{f}_{m_n(X),N_n,h_n}(x)\,dx + \left(\hat{\mu}_{m_n(X),N_n}(I^\beta) - \int_I \hat{f}_{m_n(X),N_n,h_n}(x)\,dx\right)$$

$$+\frac{1}{N_n}\sum_{i=n+L_n1}^{n+L_n+N_n} I_{\{|Y_i-m_n(X_i)|>\beta\}} + 2\cdot\sup_{t\in\mathbb{R}}\left|\hat{G}_{Y,N_n}(t)-G_Y(t)\right|$$

and

$$\mathbf{P}\{Y \in I\} \geq \int_I \hat{f}_{m_n(X),N_n,h_n}(x)\,dx - \left(\int_I \hat{f}_{m_n(X),N_n,h_n}(x)\,dx - \hat{\mu}_{m_n(X),N_n}(I_\beta)\right) - \frac{1}{N_n}\sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i - m_n(X_i)|>\beta\}} - 2\cdot\sup_{t\in\mathbb{R}}\left|\hat{G}_{Y,N_n}(t) - G_Y(t)\right|.$$

Proof. We have

$$\begin{aligned}
&\mathbf{P}\{Y \in I\}\\
&= (\mathbf{P}\{Y \in I\} - \hat{\mu}_{Y,N_n}(I)) + \left(\hat{\mu}_{Y,N_n}(I) - \hat{\mu}_{m_n(X),N_n}(I^\beta)\right)\\
&\quad + \left(\hat{\mu}_{m_n(X),N_n}(I^\beta)) - \int_I \hat{f}_{m_n(X),N_n,h_n}(x)\,dx\right)\\
&\quad + \int_I \hat{f}_{m_n(X),N_n,h_n}(x)\,dx,
\end{aligned}$$

hence concerning the first inequality it suffices to show

$$\mathbf{P}\{Y \in I\} - \hat{\mu}_{Y,N_n}(I) \leq 2\cdot\sup_{t\in\mathbb{R}}\left|\hat{G}_{Y,N_n}(t) - G_Y(t)\right| \tag{9}$$

and

$$\hat{\mu}_{Y,N_n}(I) - \hat{\mu}_{m_n(X),N_n}(I^\beta) \leq \frac{1}{N_n}\sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i - m_n(X_i)|>\beta\}}. \tag{10}$$

In case $I = [a, b]$ for some $a, b \in \mathbb{R}$ we have

$$\begin{aligned}
&\mathbf{P}\{Y \in I\} - \hat{\mu}_{Y,N_n}(I)\\
&= \mathbf{P}\{Y \in (-\infty, b]\} - \mathbf{P}\{Y \in (-\infty, a)\} - \hat{\mu}_{Y,N_n}((-\infty, b]) + \hat{\mu}_{Y,N_n}((-\infty, a))\\
&\leq |\mathbf{P}\{Y \in (-\infty, b]\} - \hat{\mu}_{Y,N_n}((-\infty, b])| + |\mathbf{P}\{Y \in (-\infty, a)\} - \hat{\mu}_{Y,N_n}((-\infty, a))|\\
&\leq 2\cdot\sup_{t\in\mathbb{R}}\left|\hat{G}_{Y,N_n}(t) - G_Y(t)\right|
\end{aligned}$$

by the definition of the (empirical) cdf. and the continuity of the measure from below. In the same way we get (9) in three other cases $I = (a, b]$,$I = [a, b)$ or $I = (a, b)$ for some $a, b \in \mathbb{R}$.

In order to prove (10) we observe that $x \in I$ and $|x - z| \leq \beta$ implies $z \in I^\beta$, hence

$$\hat{\mu}_{Y,N_n}(I)$$

$$\leq \frac{1}{N_n}\sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{Y_i\in I, |Y_i-m_n(X_i)|\leq\beta\}} + \frac{1}{N_n}\sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i-m_n(X_i)|>\beta\}}$$

$$\leq \frac{1}{N_n}\sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{m_n(X_i)\in I^\beta\}} + \frac{1}{N_n}\sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i-m_n(X_i)|>\beta\}}$$

$$= \hat{\mu}_{m_n(X),N_n}(I^\beta) + \frac{1}{N_n}\sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i-m_n(X_i)|>\beta\}}.$$

Concerning the second part of the assertion of Lemma 3 we observe

$$\begin{aligned}
&\mathbf{P}\{Y\in I\}\\
&= (\mathbf{P}\{Y\in I\} - \hat{\mu}_{Y,N_n}(I)) + \left(\hat{\mu}_{Y,N_n}(I) - \hat{\mu}_{m_n(X),N_n}(I_\beta)\right)\\
&\quad + \left(\hat{\mu}_{m_n(X),N_n}(I_\beta) - \int_I \hat{f}_{m_n(X),N_n,h_n}(x)\,dx\right)\\
&\quad + \int_I \hat{f}_{m_n(X),N_n,h_n}(x)\,dx,
\end{aligned}$$

hence to prove the second inequality it suffices to show

$$\mathbf{P}\{Y\in I\} - \hat{\mu}_{Y,N_n}(I) \geq -2\cdot\sup_{t\in\mathbb{R}}\left|\hat{G}_{Y,N_n}(t) - G_Y(t)\right| \tag{11}$$

and

$$\hat{\mu}_{Y,N_n}(I) - \hat{\mu}_{m_n(X),N_n}(I_\beta) \geq -\frac{1}{N_n}\sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i-m_n(X_i)|>\beta\}}. \tag{12}$$

By mimicking the proof of (9) and (10) we can show (11) and (12), which completes the proof. □

Proof of Theorem 1. Let I be an arbitrary interval of length $|I| > \kappa_n$. Then

$$\begin{aligned}
&\int_I \hat{g}_n^{(upper)}(x)\,dx\\
&\geq \int_I \hat{f}_{m_n(X),N_n,h_{N_n}}(x)\,dx + \left(\epsilon_n + \gamma_n + \frac{2\cdot\sqrt{\log N_n}}{\sqrt{N_n}}\right)\\
&\quad + \int_I \sup_{\substack{J\, interval,\, x\in J\\ |J|>\kappa_n}} \frac{\hat{\mu}_{m_n(X),N_n}(J^{\hat{\beta}}) - \int_J \hat{f}_{m_n(X),N_n,h_{N_n}}(t)\;dt}{\kappa_n}\,dx\\
&\geq \int_I \hat{f}_{m_n(X),N_n,h_{N_n}}(x)\,dx + \left(\epsilon_n + \gamma_n + \frac{2\cdot\sqrt{\log N_n}}{\sqrt{N_n}}\right)
\end{aligned}$$

$$+\hat{\mu}_{m_n(X),N_n}(I^{\hat{\beta}}) - \int_I \hat{f}_{m_n(X),N_n,h_{N_n}}(t)\, dt$$

and

$$\int_I \hat{g}_n^{(lower)}(x)\, dx$$
$$\leq \int_I \hat{f}_{m_n(X),N_n,h_{N_n}}(x)\, dx - \left(\epsilon_n + \gamma_n + \frac{2 \cdot \sqrt{\log N_n}}{\sqrt{N_n}}\right)$$
$$- \int_I \sup_{\substack{J\ interval,\, x \in J \\ |J| > \kappa_n}} \frac{\int_J \hat{f}_{m_n(X),N_n,h_{N_n}}(t)\, dt - \hat{\mu}_{m_n(X),N_n}(J_{\hat{\beta}})}{\kappa_n}\, dx$$
$$\leq \int_I \hat{f}_{m_n(X),N_n,h_{N_n}}(x)\, dx - \left(\epsilon_n + \gamma_n + \frac{2 \cdot \sqrt{\log N_n}}{\sqrt{N_n}}\right)$$
$$- \left(\int_I \hat{f}_{m_n(X),N_n,h_{N_n}}(t)\, dt - \hat{\mu}_{m_n(X),N_n}(I_{\hat{\beta}})\right).$$

Hence Lemma 3 implies that it suffices to show that we have outside of an event, whose probability is bounded from above by δ, that the following inequalities hold for all intervals I of length $|I| > \kappa_n$:

$$2 \cdot \sup_{t \in \mathbb{R}} \left| \hat{G}_{Y,N_n}(t) - G_Y(t) \right| \leq \frac{2 \cdot \sqrt{\log N_n}}{\sqrt{N_n}} \tag{13}$$

and

$$\frac{1}{N_n} \sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i - m_n(X_i)| > \beta\}} \leq \epsilon_n + \gamma_n. \tag{14}$$

Using the Dvoretzky-Kiefer-Wolfowitz inequality (cf., Massart (1990)) we can show that probability that (13) does not hold is bounded from above by

$$2 \cdot \exp\left(-2 \cdot N_n \cdot \frac{\log N_n}{N_n}\right) = \frac{2}{N^2}.$$

Together with Lemma 1 this implies that probability that (13) or (14) does not hold is bounded from above by δ. □

5.2 Proof of Theorem 2

In the proof we will need the following Lemma from Bott, Felber and Kohler (2015).

Lemma 4 *Let the kernel function K be a symmetric, bounded density which is monotonically decreasing on $\mathbb{R}_+$ and let $h_n > 0$. Then it holds*

$$\int \left| K\left(\frac{y-z_1}{h_n}\right) - K\left(\frac{y-z_2}{h_n}\right)\right| dy \leq 2 \cdot K(0) \cdot |z_1 - z_2|$$

for arbitrary $z_1, z_2 \in \mathbb{R}$.

Proof. See Lemma 4.1 in Bott, Felber and Kohler (2015). □

Proof of Theorem 2. Since

$$\begin{aligned}
&\left|\hat{g}_n^{(upper)}(y) - \hat{g}_n^{(lower)}(y)\right| \\
&= \frac{2}{\kappa_n} \cdot \left(\epsilon_n + \gamma_n + \frac{2 \cdot \sqrt{\log N_n}}{\sqrt{N_n}}\right) \\
&\quad + \sup_{\substack{J\, interval,\, y \in J \\ |J| > \kappa_n}} \frac{\hat{\mu}_{m_n(X),N_n}(J^{\hat{\beta}}) - \int_J \hat{f}_{m_n(X),N_n,h_{N_n}}(t)\, dt}{\kappa_n} \\
&\quad + \sup_{\substack{J\, interval,\, y \in J \\ |J| > \kappa_n}} \frac{\int_J \hat{f}_{m_n(X),N_n,h_{N_n}}(t)\, dt - \hat{\mu}_{m_n(X),N_n}(J_{\hat{\beta}})}{\kappa_n} \\
&\leq \frac{2}{\kappa_n} \cdot \left(\epsilon_n + \gamma_n + \frac{2 \cdot \sqrt{\log N_n}}{\sqrt{N_n}}\right) \\
&\quad + \sup_{\substack{J\, interval, \\ |J| > \kappa_n}} \frac{\hat{\mu}_{m_n(X),N_n}(J^{\hat{\beta}}) - \int_J \hat{f}_{m_n(X),N_n,h_{N_n}}(t)\, dt}{\kappa_n} \\
&\quad + \sup_{\substack{J\, interval, \\ |J| > \kappa_n}} \frac{\int_J \hat{f}_{m_n(X),N_n,h_{N_n}}(t)\, dt - \hat{\mu}_{m_n(X),N_n}(J_{\hat{\beta}})}{\kappa_n},
\end{aligned}$$

the definitions of ϵ_n and γ_n (from which we conclude $\gamma_n \leq \sqrt{(\log N_n)/N_n}$ for n large) imply that it suffices to show that outside of an event, whose probability tends to δ for $n \to \infty$ we have

$$\begin{aligned}
&\sup_{J\, interval,\, |J| > \kappa_n} \left(\hat{\mu}_{m_n(X),N_n}(J^{\hat{\beta}}) - \int_J \hat{f}_{m_n(X),N_n,h_{N_n}}(t)\, dt\right) \qquad (15) \\
&\leq \frac{c_1}{2} \cdot \left(h_{N_n}^p + \frac{1}{\sqrt{N_n \cdot h_{N_n}}}\right) + 3 \cdot \frac{\log n}{n} + 6 \cdot \frac{\sqrt{\log N_n}}{\sqrt{N_n}} \\
&\quad + \frac{4 \cdot (K(0) + \|g\|_\infty) \cdot \hat{\beta}}{\min\{h_{N_n}, 1\}}
\end{aligned}$$

and

$$\sup_{J\, interval,\, |J|>\kappa_n} \left(\int_J \hat{f}_{m_n(X),N_n,h_{N_n}}(t)\,dt - \hat{\mu}_{m_n(X),N_n}(J_{\hat{\beta}}) \right) \tag{16}$$
$$\leq \frac{c_1}{2} \cdot \left(h_{N_n}^p + \frac{1}{\sqrt{N_n \cdot h_{N_n}}} \right) + 3 \cdot \frac{\log n}{n} + 6 \cdot \frac{\sqrt{\log N_n}}{\sqrt{N_n}}$$
$$+ \frac{4 \cdot (K(0) + \|g\|_\infty) \cdot \hat{\beta}}{\min\{h_{N_n}, 1\}}.$$

From the proof of Theorem 1 (cf., proof of (13) and (14)) we know that outside of an event, whose probability is bounded from above by δ, we have

$$\frac{1}{N_n} \sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i - m_n(X_i)| > \hat{\beta}\}} \leq \epsilon_n + \gamma_n \tag{17}$$

and

$$\sup_{t \in \mathbb{R}} \left| \hat{G}_{Y,N_n}(t) - G_Y(t) \right| \leq \frac{\sqrt{\log N_n}}{\sqrt{N_n}}. \tag{18}$$

Hence it suffices to show that on the event, that (17) and (18) hold, we have outside of an event, whose probability tends to zero for $n \to \infty$, that (15) and (16) hold. So from now on we assume that (17) and (18) hold.

In order to prove (15), let J be an arbitrary interval. Then

$$\hat{\mu}_{m_n(X),N_n}(J^{\hat{\beta}}) - \int_J \hat{f}_{m_n(X),N_n,h_{N_n}}(t)\,dt$$
$$= \hat{\mu}_{m_n(X),N_n}(J^{\hat{\beta}}) - \hat{\mu}_{Y,N_n}(J^{2\cdot\hat{\beta}})$$
$$+ \hat{\mu}_{Y,N_n}(J^{2\cdot\hat{\beta}}) - \int_{J^{2\cdot\hat{\beta}}} g(y)\,dy$$
$$+ \int_{J^{2\cdot\hat{\beta}}} g(y)\,dy - \int_J g(y)\,dy$$
$$+ \int_J g(y)\,dy - \int_J \hat{f}_{Y,N_n,h_{N_n}}(y)\,dy$$
$$+ \int_J \hat{f}_{Y,N_n,h_{N_n}}(y)\,dy - \int_J \hat{f}_{m_n(X),N_n,h_{N_n}}(y)\,dy$$
$$= \sum_{k=1}^{5} T_{k,n}.$$

As in the proof of Lemma 3 we see

$$T_{1,n} \leq \frac{1}{N_n} \sum_{i=n+L_n+1}^{n+L_n+N_n} I_{\{|Y_i - m_n(X_i)| > \hat{\beta}\}}$$

and

$$T_{2,n} \leq 2 \cdot \sup_{t \in \mathbb{R}} \left| \hat{G}_{Y,N_n}(t) - G_Y(t) \right|,$$

from which we conclude (via (17), (18) and the definitions of ϵ_n and γ_n) that we have for n large

$$T_{1,n} + T_{2,n} \leq \frac{\log n}{n} + 3 \cdot \frac{\sqrt{\log N_n}}{\sqrt{N_n}}.$$

Furthermore the boundedness of density g implies

$$T_{3,n} \leq 4 \cdot \hat{\beta} \cdot \|g\|_\infty.$$

Next we derive an upper bound on $T_{4,n}$. We have

$$\begin{aligned} T_{4,n} &\leq \int_{\mathbb{R}} |\hat{f}_{Y,N_n,h_{N_n}}(y) - g(y)| \, dy \\ &= \int_{\mathbb{R}} |\hat{f}_{Y,N_n,h_{N_n}}(y) - g(y)| \, dy - \mathbf{E} \int_{\mathbb{R}} |\hat{f}_{Y,N_n,h_{N_n}}(y) - g(y)| \, dy \\ &\quad + \mathbf{E} \int_{\mathbb{R}} |\hat{f}_{Y,N_n,h_{N_n}}(y) - g(y)| \, dy. \end{aligned}$$

By standard application of the McDiarmid's inequality (cf., e.g., Theorem A.2 in Györfi et al. (2012) and proof Theorem 1 in Devroye et al. (2012)) we see that outside of an event, whose probability tends to zero for $n \to \infty$, we have

$$\int_{\mathbb{R}} |\hat{f}_{Y,N_n,h_{N_n}}(y) - g(y)| \, dy - \mathbf{E} \int_{\mathbb{R}} |\hat{f}_{Y,N_n,h_{N_n}}(y) - g(y)| \, dy \leq \frac{\sqrt{\log N_n}}{\sqrt{N_n}}.$$

Since the supports of g and K are compact and since h_{N_n} is bounded there exists a compact set $A \subseteq \mathbb{R}$ such that

$$\mathbf{E} \int_{\mathbb{R}} |\hat{f}_{Y,N_n,h_{N_n}}(y) - g(y)| \, dy = \mathbf{E} \int_A |\hat{f}_{Y,N_n,h_{N_n}}(y) - g(y)| \, dy.$$

By standard arguments used in analysis of the L_1 error of the density estimates (cf., e.g., proof of Theorem 1 in Felber, Kohler and Krzyżak (2015a)) we can conclude from the (p, C)–smoothness of the density g and from the compactness of its support that the latter term can be bounded from above by

$$\frac{c_1}{2} \cdot \left(h_{N_n}^p + \frac{1}{\sqrt{N_n \cdot h_{N_n}}} \right).$$

Summarizing these results we see that outside of an event, whose probability tends to zero for $n \to \infty$, we have

$$T_{4,n} \leq \frac{\sqrt{\log N_n}}{\sqrt{N_n}} + \frac{c_1}{2} \cdot \left(h_{N_n}^p + \frac{1}{\sqrt{N_n \cdot h_{N_n}}} \right).$$

In order to bound $T_{5,n}$, we observe

$$\begin{aligned} T_{5,n} &= \frac{1}{N_n \cdot h_{N_n}} \cdot \sum_{i=n+L_n+1}^{n+L_n+N_n} \int_J \left(K\left(\frac{Y_i - y}{h_{N_n}}\right) - K\left(\frac{m_n(X_i) - y}{h_{N_n}}\right)\right) dy \\ &\leq \frac{1}{N_n} \cdot \sum_{i=n+L_n+1}^{n+L_n+N_n} 2 \cdot I_{\{|Y_i - m_n(X_i)| > \hat{\beta}\}} \\ &\quad + \frac{1}{N_n \cdot h_{N_n}} \cdot \sum_{i=n+L_n+1}^{n+L_n+N_n} 2 \cdot K(0) \cdot \hat{\beta}, \end{aligned}$$

where the last inequality follows from the fact that K is a density and from Lemma 4. Application of (17) yields for large n

$$T_{5,n} \leq 2 \cdot (\epsilon_n + \gamma_n) + 2 \cdot K(0) \cdot \frac{\hat{\beta}}{h_{N_n}} \leq 2 \cdot \frac{\log n}{n} + 2 \cdot \frac{\sqrt{\log N_n}}{\sqrt{N_n}} + \frac{2 \cdot K(0) \cdot \hat{\beta}}{h_{N_n}},$$

which implies (15).

In the same way one can prove (16), which completes the proof of Theorem 2. □

6 Acknowledgment

The first author would like to thank the German Research Foundation (DFG) for funding this project within the Collaborative Research Center 805. The second author gratefully acknowledges support from the Natural Sciences and Engineering Research Council of Canada.

References

[1] Bayarri, M. J., Berger, J. O., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J., Parthasarathy, R. J., Paulo, R., Sacks, J., and Walsh, D. (2007). Computer model validation with functional output. *The Annals of Statistics*, **35**, pp. 1874–1906.

[2] Bichon, B., Eldred, M., Swiler, M., Mahadevan, S. and McFarland, J. (2008). Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA Journal*, **46**, pp. 2459–2468.

[3] Bickel, P. J., and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *The Annals of Statistics*, **1**, pp. 1071-1095.

[4] Bott, A. K., Felber, T., and Kohler, M. (2015). Estimation of a density in a simulation model. *Journal of Nonparametric Statistics*, **27**, pp. 271-285.

[5] Bourinet, J.-M., Deheeger, F. and Lemaire, M. (2011). Assessing small failure probabilities by combined subset simulation and support vector machines. *Structural Safety*, **33**, pp. 343–353.

[6] Bucher, C. and Bourgund, U. (1990). A fast and efficient response surface approach for structural reliability problems. *Structural Safety*, **7**, pp. 57-66.

[7] Coombs, Clyde H. (1964). *A Theory of Data.* New York, John Wiley & Sons.

[8] Das, P.-K. and Zheng, Y. (2000). Cumulative formation of response surface and its use in reliability analysis. *Probabilistic Engineering Mechanics*, **15**, pp. 309-315.

[9] Deheeger, F. and Lemaire, M. (2010). Support vector machines for efficient subset simulations: ^{2}SMART method. In: *Proceedings of the 10th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP10)*, Tokyo, Japan.

[10] Devroye, L., Felber, T., and Kohler, M. (2013). Estimation of a density using real and artificial data. *IEEE Transactions on Information Theory*, **59**, No. 3, pp. 1917-1928.

[11] Devroye, L., Felber, T., Kohler, M., and Krzyżak, A. (2012). L_1–consistent estimation of the density of residuals in random design regression models. *Statistics and Probability Letters*, **82**, pp. 173-179.

[12] Felber, T., Kohler, M., and Krzyżak, A. (2015a). Adaptive density estimation based on real and artificial data. *Journal of Nonparametric Statistics*, **27**, pp. 1-18.

[13] Felber, T., Kohler, M., and Krzyżak, A. (2015b). Density estimation with small measurement errors. *IEEE Transactions on Information Theory*, **61**, pp. 3446-3456.

[14] Giné, E., and Nickl, R. (2010). Confidence bands in density estimation. *The Annals of Statistics*, **38**, pp. 1122-1170.

[15] Goh, J., Bingham, D., Holloway, J. P., Grosskopf, M. J., Kuranz, C. C., and Rutter, E. (2013). Prediction and computer model calibration using outputs from mulitfidelity simulators. *Technometrics*, **55**, pp. 501-512.

[16] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). A Distribution-Free Theory of Nonparametric Regression. *Springer-Verlag*, New York.

[17] Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *The Annals of Statistics*, **20**, pp. 675-694.

[18] Hall, P. , and Horowitz, J. (2013). A simple bootstrap method for constructing nonparametric confidence bands for functions. *The Annals of Statistics*, **41**, pp. 1892-1921.

[19] Hall, P. , and Titterington, D. M. (1988). On confidence bands in nonparametric density estimation and regression. *Journal of Multivariate Analysis*, **27**, pp. 228-254.

[20] Han, G., Santner, T. J., Rawlinson, J. J. (2009). Simultaneous determination of tuning and calibration parameters for computer experiments. *Technometrics*, **51**, pp. 464-474.

[21] Higdon, D., Gattiker, J., Lawrence, E., Jackson, C., Tobis, M., Pratola, M., Habib, S., Heitmann, K., and Price, S. (2013). Computer model calibration using the ensemble kalman filter. *Technometrics*, **55**, pp. 488–500.

[22] Hurtado, J. (2004). *Structural Reliability – Statistical Learning Perspectives.* Vol. 17 of lecture notes in applied and computational mechanics. Springer.

[23] Kaymaz, I. (2005). Application of Kriging method to structural reliability problems. *Strutural Safety*, **27**, pp. 133–151.

[24] Kennedy, M. C., and O'Hagan, A. (2001). Bayesian calibration of computer models (with discussion). *Journal of the Royal Statistical Society: Series B*, **63**, pp. 425-464.

[25] Kim, S.-H. and Na, S.-W. (1997). Response surface method using vector projected sampling points. *Structural Safety*, **19**, pp. 3–19.

[26] Kohler, M., Krzyżak, A., Mallapur, S., and Platz, R. (2017). Uncertainty Quantification in Case of Imperfect Models: A Non-Bayesian Approach. Submitted for publication.

[27] Mallapur, S., and Platz, R. (2016). Quantification and Evaluation of Uncertainty in the Mathematical Modelling of a Suspension Strut using Bayesian Model Validation Approach. Submitted for publication.

[28] Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Annals of Probability* **18**, pp. 1269–1283.

[29] Papadrakakis, M. and Lagaros, N. (2002). Reliability–based structural optimization using neural networks and Monte Carlo simulation. *Computer Methods in Applied Mechanics and Engineering*, **191**, pp. 3491–3507.

[30] Santner, T. J., Williams, B. J. and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments.* Springer-Verlag, New York.

[31] Tuo, R., and Wu, C. F. J. (2016). Efficient calibration for imperfect computer models. To appear in *Annals of Statistics.*

[32] Wang, S., Chen, W., and Tsui, K. L. (2009). Bayesian validation of computer models. *Technometrics*, **51**, pp. 439-451.