Felix Wittich*, Lars Kistner, Andreas Kroll, Christopher Schott, and Thomas Niendorf

On data-driven nonlinear uncertainty modeling: Methods and application for control-oriented surface condition prediction in hard turning

Zur nichtlinearen datengetriebenen Unsicherheitsmodellierung: Methoden und Anwendung für die regelungsorientierte Randschichteigenschaftsprädiktion beim Hartdrehen

https://doi.org/10.1515/teme-2020-0057
Received July 20, 2020; accepted September 8, 2020

Abstract: In this article, two data-driven modeling approaches are investigated, which allow an explicit modeling of uncertainty. For this purpose, parametric Takagi-Sugeno multi-models with bounded-error parameter estimation and nonparametric Gaussian process regression are applied and compared. These models can for instance be used for robust model-based control design. As an application, the prediction of residual stresses during hard turning depending on the machining parameters and the initial hardness is considered.

Keywords: Uncertainty, Takagi Sugeno model, Gaussian process regression, hard turning.

Zusammenfassung: In diesem Artikel werden zwei datengetriebene Modellierungsansätze für den Einsatz bei zerspanenden Prozessen untersucht, die eine explizite Modellierung der Unsicherheit erlauben. Dazu werden ein parametrisches Takagi-Sugeno-Multi-Modell mit einer Bounded-Error Parameterschätzung und eine nichtparametrische Gaussprozessregression eingesetzt und verglichen. Diese Modelle können beispielsweise für den robusten Vorsteuerungs- und Regelungsentwurf eingesetzt werden. Als Anwendung wird die Prädiktion der Eigenspannungen beim Hartdrehen in Abhängigkeit der Maschinenstellgrößen und der initialen Härte betrachtet.

Schlagwörter: Unsicherheit, Takagi-Sugeno-Modelle, Gaußprozessregression, Hartdrehen.

***Corresponding author: Felix Wittich,** Department of Measurement and Control, Faculty of Mechanical Engineering, University of Kassel, 34125 Kassel, Germany, e-mail: felix.wittich@mrt.uni-kassel.de
Lars Kistner, Andreas Kroll, Department of Measurement and Control, Faculty of Mechanical Engineering, University of Kassel, 34125 Kassel, Germany
Christopher Schott, Thomas Niendorf, Institute of Materials Engineering, Metallic Materials, University of Kassel, 34125 Kassel, Germany

1 Introduction

Because machining affects the surface condition of a workpiece, it is of great interest to be able to control these processes. As a final machining step, hard turning offers the possibility to directly influence the surface condition and therefore to replace subsequent finishing steps like grinding. Important surface integrity parameters are the near-surface residual stresses as they have a major impact on structural integrity [7]. Measurement of the surface residual stress condition is usually conducted after manufacturing by X-ray diffractometry resulting in a very expensive and time consuming procedure. The availability of real-time residual stress predictions would permit to automatically adjust machining parameters during manufacturing to establish the required component properties in spite of process disturbances and parameter variation.

Finite element simulations are time consuming and therefore not suitable for real-time prediction [1]. Analytical approaches are difficult to establish as specific in-depth knowledge of the material and the process is needed [21]. Data-driven approaches allow for both high approximation accuracies and real-time capability. However, to use empirical models effectively, it is important to asses their trustworthiness. Often little attention is paid to the quantification of the model uncertainty in empirical modeling for machining processes. Therefore, this paper focuses on approaches that quantify the uncertainty. The obtained models can e. g. be used for robust model-based control of the final residual stress state in hard turning processes. The approaches can also easily be extended to other machining processes and materials.

The rest of the paper is organized as follows: In section 2 the problem of modeling with uncertainty quantification is presented and related work is reviewed. Section 3 and 4 introduce the considered modeling approaches. Section 5 presents the case study and in Section 6 a conclusion and an outlook is given.

2 Problem statement and related work

In general, the considered problem is finding the unknown nonlinear functional relationship $f : \mathbb{R}^n \rightarrow \mathbb{R}$ between n input variables $\boldsymbol{x} \in \mathbb{R}^n$ and the output variable $y \in \mathbb{R}$

$$y(k) = f(\boldsymbol{x}(k)) + e(k), k = 1, \ldots, N \tag{1}$$

with the error term $e(k)$ and based on a data set $Z^N = \{\boldsymbol{x}(k), y(k)\}_{k=1}^N$ with N elements, containing measurements of the inputs and the output of the process under consideration. In the probabilistic setting the error is described by a probability distribution. Any arbitrary probability distribution can be assumed but the most common assumption is a normally distributed error. In general, the error $\boldsymbol{e} = [e(1) \ldots e(N)]^T$ is multivariate normally distributed:

$$\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{\mu}_\mathrm{e}, \boldsymbol{\Sigma}_\mathrm{e}). \tag{2}$$

with the vector of means $\boldsymbol{\mu}_\mathrm{e} = [\mu_{\mathrm{e},1} \ldots \mu_{\mathrm{e},N}]^T \in \mathbb{R}^N$ and the covariance matrix $\boldsymbol{\Sigma}_\mathrm{e} \in \mathbb{R}^{N \times N}$, $\Sigma_{\mathrm{e},i,j} = \mathrm{cov}(e(i), e(j))$, where $cov(x, y)$ gives the covariance between the two inputs $e(i)$ and $e(j)$. An alternative error description, which is independent of probabilistic assumptions, is the bounded-error (BE) approach. Here, the assumption is made that the prediction error $e(k)$ lies in an interval with defined boundaries:

$$e(k)_\mathrm{min} \leq e(k) \leq e(k)_\mathrm{max}, k = 1, \ldots, N. \tag{3}$$

In this paper BE methods are used for parameter estimation of non-linear Takagi-Sugeno (TS) models. Parametric TS models are well suited to model nonlinear dependencies with a compact model description that can be used for model-based control design. In a previous work [23], TS models with least squares based parameter estimation outperformed a multiple linear regression approach in predicting residual stresses in hard turning. Only few publications report on TS models with BE parameter estimation. In [11], TS models with BE parameter estimation were used for fault detection for a valve in a sugar production process. In [15], the BE approach was used with TS models for state estimation in context of actuator fault detection.

Nonparametric Gaussian process regression (GPR) models were established in [17] in a machine learning context and are especially suitable for small data sets. They allow modeling of uncertainty and are considered as an alternative to the set-based TS modeling in this paper. However, as a nonparametric approach they are not suitable for model-based control schemes. Only few publications exist where GPR is used for empirical modeling of machining processes. In [10], tool wear in a dry turning process is predicted from cutting force measurements. In [24] GPR was used for modeling the surface roughness in end face milling.

Several publications with different empirical modeling approaches for prediction of the surface integrity in hard turning have been published so far. In [4] polynomial regression combined with the response surface design methodology is proposed to model residual stress profiles for different materials in turning operations. Artificial neural networks (ANN) were used in [25] to predict the longitudinal and circumferential residual stresses at five different depths.

To further explore this topic, in the present paper the influences of different initial hardness values in combination with selected feed rates and cutting speeds on the residual stress profiles are modeled. The reviewed publications do not explicitly consider uncertainty. The proposed approaches in the present work allow quantifying the reliability of the model. Therefore, they are the foundation for a robust model-based control scheme of the relevant surface condition in hard turning. Furthermore, both approaches are in particular suitable for modeling with sparse data sets. Other approaches such as ANNs require larger data sets for training. Usually, BE estimation is used for small modeling problems as it is computationally expensive. In this contribution it will be applied for TS models with a significant number of parameters.

3 Parametric bounded-error identification of Takagi-Sugeno models

The first considered approach for the identification of prediction models is to use BE parameter estimation for parametric Takagi-Sugeno models such that the uncertainty is described by guaranteed error bounds. This leads to a set of feasible parameters for the model resulting from the error bounds.

3.1 Takagi-Sugeno models

TS fuzzy models [20] consist of $c \in \mathbb{N}_+$ superposed local models $\hat{y}_j = f(\boldsymbol{\theta}_{j,\mathrm{LM}}, \tilde{\boldsymbol{\varphi}}) : \mathbb{R}^{n_p} \rightarrow \mathbb{R}$, weighted by their corresponding fuzzy basis function $\phi_j(\boldsymbol{z}) : \mathbb{R}^{n_z} \rightarrow [0, 1]$, which

depends on the scheduling variable $\boldsymbol{z} = [z_1 \dots z_{n_z}]^\top \in \mathbb{R}^{n_z}$:

$$\hat{y}(\boldsymbol{z}, \boldsymbol{\theta}_{j,\mathrm{LM}}, \tilde{\boldsymbol{\varphi}}) = \sum_{j=1}^{c} \phi_j(\boldsymbol{z}) \cdot \hat{y}_j(\boldsymbol{\theta}_{j,\mathrm{LM}}, \tilde{\boldsymbol{\varphi}}). \tag{4}$$

Locally affine models are used:

$$\hat{y}_j(\boldsymbol{\theta}_{j,\mathrm{LM}}, \tilde{\boldsymbol{\varphi}}) = \sum_{r=0}^{n} \theta_{j,r,\mathrm{LM}} \cdot \tilde{\varphi}_r = \boldsymbol{\theta}_{j,\mathrm{LM}} \cdot \tilde{\boldsymbol{\varphi}}, \tag{5}$$

with $\tilde{\varphi}_r$ being the r-th element of the vector

$$\tilde{\boldsymbol{\varphi}} = [1\ x_1 \dots x_{n_p}]^\top \tag{6}$$

and $\theta_{j,r,\mathrm{LM}}$ being the r-th element of the corresponding local parameter vector $\boldsymbol{\theta}_{j,\mathrm{LM}} \in \mathbb{R}^n$. The TS model (4) can be rewritten as:

$$\hat{y} = \bar{\boldsymbol{\varphi}}^T \boldsymbol{\theta}_{\mathbf{LM}} \tag{7}$$

with the extended regression vector

$$\bar{\varphi} = \left[\phi_1\ \phi_1 x_1\ \dots\ \phi_1 x_{n_p} | \dots | \phi_c\ \phi_c x_1\ \dots\ \phi_c x_{n_p}\right]^T \tag{8}$$

and the vector of local model parameters

$$\boldsymbol{\theta}_{\mathbf{LM}} = \left[a_{0,1}\ a_{1,1}\ \dots\ a_{n_p,1} | \dots | a_{0,c}\ a_{1,c}\ \dots\ a_{n_p,c}\right]^T. \tag{9}$$

The fuzzy basis function $\phi_j(\boldsymbol{z})$ defines the validity region for the corresponding j-th local model. It is given by

$$\phi_j(\boldsymbol{z}) = \frac{\mu_j(\boldsymbol{z})}{\sum_{m=1}^{c} \mu_m(\boldsymbol{z})}, \tag{10}$$

with the membership function $\mu_j(\boldsymbol{z})$. In this contribution, multivariate membership functions obtained from fuzzy-c-means (FCM) clustering are used. FCM clustering converges fast and the membership functions require only few parameters while good modeling results are achieved, see e. g. [12]. The $\mu_j(\boldsymbol{z})$ are given by

$$\mu_j(\boldsymbol{z}) = \left[\sum_{i=1}^{c} \left(\frac{\|\boldsymbol{z} - \boldsymbol{v}_j\|^2_{\boldsymbol{A}_j}}{\|\boldsymbol{z} - \boldsymbol{v}_i\|^2_{\boldsymbol{A}_i}}\right)^{\frac{1}{\nu-1}}\right]^{-1} \tag{11}$$

with the norm inducing matrix $\boldsymbol{A}_j$ defining individual distance norms

$$\|\boldsymbol{z} - \boldsymbol{v}_j\|^2_{\boldsymbol{A}_j} = (\boldsymbol{z} - \boldsymbol{v}_j)^\top \boldsymbol{A}_j (\boldsymbol{z} - \boldsymbol{v}_j), \tag{12}$$

the fuzziness parameter $\nu \in \mathbb{R}^{>1}$, and the partition's prototypes $\boldsymbol{v}_j, \boldsymbol{v}_i \in \mathbb{R}^{n_z}$ aggregated in the parameter vector $\boldsymbol{\theta}_{\mathrm{MF}}$, such that $\mu_j(\boldsymbol{z}) = \mu_j(\boldsymbol{\theta}_{\mathrm{MF}}, \boldsymbol{z})$. For simplicity, in this contribution, $\boldsymbol{A}_j = \boldsymbol{I}\ \forall j$ with the identity matrix $\boldsymbol{I}$ is chosen, inducing the Euclidean norm. The membership functions (11) fulfill the orthogonality condition

$$\sum_{j=1}^{c} \mu_j(\boldsymbol{z}) = 1\ \forall\ \boldsymbol{z} \Longrightarrow \phi_j(\boldsymbol{z}) \equiv \mu_j(\boldsymbol{z}). \tag{13}$$

The usual optimization problem resulting from the task of point estimation and a least squares type cost function is

$$\underset{\boldsymbol{\theta}_{\mathrm{LM}}, \boldsymbol{\theta}_{\mathrm{MF}}}{\operatorname{argmin}} \sum_{k=1}^{N} \left(y(k) - \sum_{j=1}^{c} \mu_j(\boldsymbol{z}(k), \boldsymbol{\theta}_{\mathrm{MF}}) \cdot \hat{y}_j(k, \boldsymbol{\theta}_{j,\mathrm{LM}})\right)^2 \tag{14}$$

with $\boldsymbol{\theta}_{\mathrm{LM}} = \left[\boldsymbol{\theta}^T_{1,\mathrm{LM}} \dots \boldsymbol{\theta}^T_{c,\mathrm{LM}}\right]^T \in \mathbb{R}^{(n+1)\cdot c}$.

A two-step procedure for data-driven identification can be described as follows: i) the identification of the parameters of the membership functions $\boldsymbol{\theta}_{\mathrm{MF}}$ and ii) the identification of the local model parameters $\boldsymbol{\theta}_{\mathrm{LM}}$. In this contribution, i) is first solved using the FCM algorithm reducing ii) to a simple linear least-squares estimation. The obtained estimates are used to initialize a nonlinear optimization of (14) where the prototypes and local model parameters are optimized simultaneously with respect to the sum of squared prediction errors of the model. For the remainder of this paper, $\boldsymbol{z} = \boldsymbol{x}$ is chosen for simplicity.

3.2 Bounded-error parameter estimation

A set-based alternative to the aforementioned least squares based parameter estimation is BE parameter estimation. It assumes that the prediction error is bounded by an interval $e(k) \in [e(k)_{\min}, e(k)_{\max}]$. The result of the parameter estimation is not a single parameter vector but a feasible parameter set (FPS). For models that are nonlinear in the parameters (non-LiP) the FPS can be complex and non-connected. Methods like SIVIA [9] to approximate the non-LiP FPS are computationally expensive. For linear in the parameters (LiP) models, $\mathbb{S}_{\mathrm{FPS}}$ is a polytope, i. e. a bounded polyhedron, if $N \geq d_{\mathrm{ind}} > n_\theta$ holds, with d_{ind} the number of linearly independent inequalities and n_θ the number of parameters. The output error $e(k)$ can be chosen individually for each $k \in \{1, \dots, N\}$. In the following, the same symmetric error bound $\delta = (e(k)_{\min} - e(k)_{\max})/2$, $\delta \in \mathbb{R}^+$ for each k will be assumed. So for a LiP model $\hat{y} = \boldsymbol{\varphi}^T \boldsymbol{\theta}$, $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ such as (7) and N observations, each of the inequalities

$$y(k) - \delta \leq \boldsymbol{\varphi}^T(k)\boldsymbol{\theta} \leq y(k) + \delta, k = 1, \dots, N \tag{15}$$

restricts the FPS to two half-spaces in $\mathbb{R}^{n_\theta}$. In consequence, $2N$ bounds result.

Polytopes can be represented by the $\mathcal{H}$-representation, i. e. through the half-spaces bounding the polytope $\mathcal{P}$:

$$\mathcal{P} = \{\boldsymbol{\theta} \in \mathbb{R}^{n_\theta} | \boldsymbol{\Phi}\boldsymbol{\theta} \leq \boldsymbol{Y}\} \tag{16}$$

with

$$\boldsymbol{Y} = \begin{bmatrix} -y(1)+\delta \\ y(1)+\delta \\ -y(2)+\delta \\ y(2)+\delta \\ \vdots \\ -y(N)+\delta \\ y(N)+\delta \end{bmatrix}, \boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\varphi}(1)^T \\ \boldsymbol{\varphi}(1)^T \\ \boldsymbol{\varphi}(2)^T \\ \boldsymbol{\varphi}(2)^T \\ \vdots \\ \boldsymbol{\varphi}(N)^T \\ \boldsymbol{\varphi}(N)^T \end{bmatrix}, \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \vdots \\ \boldsymbol{\theta}_{n_\theta} \end{bmatrix}, \tag{17}$$

$\boldsymbol{\Phi} \in \mathbb{R}^{2N \times n_\theta}$, $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$, and $\boldsymbol{Y} \in \mathbb{R}^{2N}$.

The representation of a polytope as a set of vertices is also known as $\mathcal{V}$-representation. The conversion from $\mathcal{H}$- to $\mathcal{V}$-representation is a well-known problem in the field of computational geometry. Various algorithms have been developed to solve this problems. The methods can be divided into two categories: pivoting and incremental algorithms. Well established methods are the double description (DD) [5] and the reverse search method [2]. A detailed overview of Vertex-Enumeration procedures can be found in [6]. In this article the DD method was chosen being an established procedure, which is also capable of handling degenerate polytopes. The implementation from the Multi-Parametric Toolbox 3.0 (MPT3) [8] is used, which is an improved version of the original method [14].

4 Nonparametric probabilistic identification of Gaussian process regression models

An alternative approach to describe the functional relationship is to use Gaussian processes (GP). Contrary to the parametric approach in Section 3, where the goal is to infer the set of feasible parameters for the function f given the data Z^N, the distribution $p(f|Z^N)$ is now inferred directly.

A detailed explanation and derivation of GPR can be found in [18]. The idea behind a GP is to define a probability distribution over functions. In order to define a distribution over functions it is sufficient to define a distribution over the function's value at a finite set of points $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(N)$. The GP assumes that the distribution $p(f(\boldsymbol{x}(1)), \ldots, f(\boldsymbol{x}(N))), \boldsymbol{x} \in \mathbb{R}^n$ is jointly Gaussian. For regression, the prior of the function to be learnt is defined as:

$$f(\boldsymbol{x}) = \mathrm{GP}(m(\boldsymbol{x}), \kappa(\boldsymbol{x}, \boldsymbol{x}')) \tag{18}$$

where $m(\boldsymbol{x})$ is the mean function and $\kappa(\boldsymbol{x}, \boldsymbol{x}')$ is the covariance function for two points $\boldsymbol{x}$ and $\boldsymbol{x}'$ from the input space:

$$m(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})] \tag{19}$$

$$\kappa(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}[(f(\boldsymbol{x}) - m(\boldsymbol{x}))(f(\boldsymbol{x}') - m(\boldsymbol{x}'))^\top] \tag{20}$$

with a positive definite kernel $\kappa(\boldsymbol{x}, \boldsymbol{x}')$. The kernel plays a key role in GPs as it describes the dependencies between the data points. The basic idea behind all kernel functions for the covariance functions is that if two points $\boldsymbol{x}$ and $\boldsymbol{x}'$ are similar, the outputs are also similar. The choice of the kernel strongly influences the model's behavior. The GP defines a joint Gaussian distribution for a finite set of points:

$$p(y|\boldsymbol{X}) = \mathcal{N}(y|\boldsymbol{m}, \boldsymbol{K}) \tag{21}$$

with the design matrix $\boldsymbol{X} \in \mathbb{R}^{N \times n}$ describing the input data, the covariance matrix $\boldsymbol{K} \in \mathbb{R}^{N \times N}$, $K_{i,j} = \kappa(\boldsymbol{x}(i), \boldsymbol{x}(j))$, and the vector of means $\boldsymbol{m} = [m(\boldsymbol{x}(1)) \ldots m(\boldsymbol{x}(N))]^T \in \mathbb{R}^N$.

In this paper, common choices for the kernel function are considered. Firstly, the squared exponential (SE) kernel

$$\kappa_{\mathrm{SE}}(\boldsymbol{x}, \boldsymbol{x}') = \theta_{\mathrm{f,SE}}^2 \exp\left(-\frac{1}{2} d_{\mathrm{SE}}\right) \tag{22}$$

with

$$d_{\mathrm{SE}} = \frac{\sum_{i=1}^n (x_i - x_i')^2}{\theta_{\mathrm{l,SE}}^2} \tag{23}$$

and the scaling parameters $\theta_{\mathrm{f,SE}}$ and $\theta_{\mathrm{l,SE}}$ is used. Secondly, the Matern 3\2 (M3\2) kernel

$$\kappa_{\mathrm{M32}}(\boldsymbol{x}, \boldsymbol{x}') = \theta_{\mathrm{f,M32}} \left(1 + \sqrt{3} d_{\mathrm{M32}}\right) \exp\left(-\sqrt{3} d_{\mathrm{M32}}\right) \tag{24}$$

with

$$d_{\mathrm{M32}} = \frac{\sqrt{\sum_{i=1}^n (x_i - x_i')^2}}{\theta_{\mathrm{l,M32}}} \tag{25}$$

and the scaling parameters $\theta_{\mathrm{f,M32}}$ and $\theta_{\mathrm{l,M32}}$ is used.

In the multivariate case, above mentioned kernel functions have the same length scale for each variable. An automatic relevance determination (ARD) kernel modification is used to assign individual parameters for the length scale to each variable. The ARD modification $\kappa_{\mathrm{SE+ARD}}(\boldsymbol{x}, \boldsymbol{x}')$ of (22) with

$$\tilde{d}_{\mathrm{SE}} = \sum_{i=1}^n \frac{(x_i - x_i')^2}{\theta_{\mathrm{l,SE+ARD},i}^2} \tag{26}$$

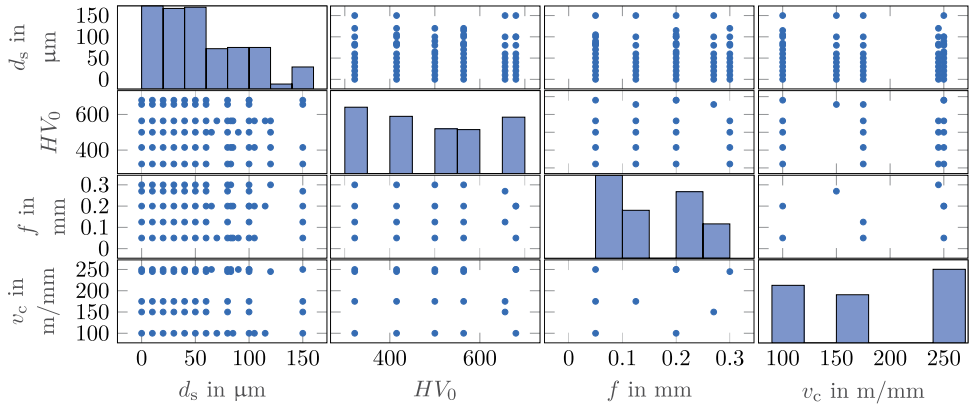


Figure 1: Matrix plot of hard turning input data with frequency distribution on the main diagonal.

and the individual length scales $\theta^2_{\mathrm{l,SE+ARD},i}$ is considered here. Also the ARD modification $\kappa_{\mathrm{M32+ARD}}(\boldsymbol{x},\boldsymbol{x}')$ of (24) with

$$\tilde{d}_{\mathrm{M32}} = \sqrt{\sum_{i=1}^{n} \frac{(x_i - x_i')^2}{\theta^2_{\mathrm{l,SE+ARD},i}}} \tag{27}$$

and the individual length scales $\theta^2_{\mathrm{l,M32+ARD},i}$ is used.

To predict the output y_* for a new test input $\boldsymbol{x}_*$, the posterior conditional probability is computed by conditioning the multivariate Gaussian (zero mean and no noise are assumed for simplicity):

$$p(y_*|\boldsymbol{x}_*,\boldsymbol{X},\boldsymbol{y}) = \mathcal{N}(y_*|m_*,\kappa_*) \tag{28}$$

$$\text{with } m_* = \boldsymbol{K}_*^T\boldsymbol{K}^{-1}\boldsymbol{y}, \tag{29}$$

$$\text{and } \kappa_* = \boldsymbol{K}_{**} - \boldsymbol{K}_*^T\boldsymbol{K}^{-1}\boldsymbol{K}_* \tag{30}$$

with the new test input $\boldsymbol{x}_* \in \mathbb{R}^n$, the covariance matrices $\boldsymbol{K} = \kappa(\boldsymbol{X},\boldsymbol{X}) \in \mathbb{R}^{N\times N}$, $\boldsymbol{K}_* = \kappa(\boldsymbol{X},\boldsymbol{X}_*) \in \mathbb{R}^N$, $\boldsymbol{K}_{**} = \kappa(\boldsymbol{X}_*,\boldsymbol{X}_*) \in \mathbb{R}$ and the vector of outputs $\boldsymbol{y} \in \mathbb{R}^N$. Predictions for the output can then be computed from the posterior mean function (29) and predictions for uncertainty from the posterior variance function (30).

Usually, it is assumed that the observed output is affected by independent, normally distributed noise $y = f(\boldsymbol{x}) + e$, $e \sim \mathcal{N}(0,\sigma_y^2)$. The noise component is then added to the covariance matrices $\boldsymbol{K}_y = \boldsymbol{K} + \sigma_y^2\boldsymbol{I}_N$. The GPR model is trained on the data set by optimization of its hyperparameters, i. e. the kernel function parameters and the noise variance. The log marginal likelihood

$$\log p(\boldsymbol{y}|\boldsymbol{X}) = -\frac{1}{2}\boldsymbol{y}\boldsymbol{K}_y^{-1}\boldsymbol{y} - \frac{1}{2}\log|\boldsymbol{K}_y| - \frac{N}{2}\log(2\pi) \tag{31}$$

is maximized using a gradient-based optimizer. A detailed derivation of the marginal likelihood function

$$p(\boldsymbol{y}|\boldsymbol{X}) = \int p(\boldsymbol{y}|\boldsymbol{f},\boldsymbol{X})p(\boldsymbol{f}|\boldsymbol{X})d\boldsymbol{f} \tag{32}$$

with $\boldsymbol{f} = [f(\boldsymbol{x}(1)),\ldots,f(\boldsymbol{x}(N))]^\top$ can be found in [18]. The log marginal likelihood's first term accounts for the data fit and the second term for model complexity. In consequence, a bias-variance trade-off is performed avoiding overfitting.

5 Case study

Hard turning is a machining process for hardened workpieces with a geometrically defined cutting edge. Hard turning not only determines the final geometry of the component but also its surface layer properties, which in turn have an effect on the components properties such as fatigue [19]. Important surface layer properties are surface residual stresses, which can be influenced by manipulating the machining parameters. Therefore, in this section the methods from section 3 and 4 will be used to derive models that predict tangential residual stress profiles σ_t from machining parameters. Furthermore, the axial residual stress σ_a and the integral widths IW_t and IW_a were determined and modeled. For the sake of brevity, only modeling of σ_t as an important indicator for surface integrity is presented. Data from hard turning experiments and subsequent X-ray diffraction measurements are used. As inputs, the feed rate f, the cutting speed v_c as well as the initial hardness of the workpiece HV_0 are considered. To further obtain the depth distribution of the residual stresses, the distance from the workpiece's surface d_s was included as explanatory variable.

5.1 Data base

The data shown in Fig. 1 stem from an experimental test campaign [13]. The cyclindrical specimens made of AISI 6150 had six different initial hardness levels $HV_0 \in \{322,$

425, 500, 564, 656, 680} HV10 (HV10 Vickers hardness). After heat treatment and prior to turning, all specimens were machined to a diameter of $d = 34\,\text{mm}$ and a length of $l = 200\,\text{mm}$. The turning operations were conducted at the Institute of Machining Technology (TU Dortmund) using a NC-controlled lathe (Monforts RNC602) using polycrystalline boron nitride (PCBN) tool inserts with a chamferwidth of $b_\gamma = 0.15\,\text{mm}$ and an angle of $\gamma_\text{f} = 25°$. The feed rate and cutting speed were varied in the intervals $f \in [0.05, 0.3]\,\text{mm}$ and $v_\text{c} \in [100, 250]\,\text{m/min}$. The residual stresses were measured using a Ψ-diffractometer using CrKα radiation. The measurements were taken for 10 different distances from the specimens' surface in the interval $d_\text{s} \in [0, 150]\,\mu\text{m}$ by electrolytically removing material. A total of $N = 324$ data points were generated. This results in a multi-input-single-output regression problem with a four dimensional input variable space consisting of $d_\text{s}, f, v_\text{c}$ and HV_0 used to predict the residual stress σ_t. The data were standardized to have zero mean and unit standard deviation. However, for ease of interpretation all figures and tables refer to non-standardized data.

5.2 Model validation

To evaluate and compare the predictive performance of the trained models, quantitative accuracy metrics are considered. Two metrics commonly used for regression problems, the root mean squared error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(y(k) - \hat{y}(k))^2} \tag{33}$$

and the coefficient of determination R^2

$$R^2 = 1 - \frac{\sum_{k=1}^{N}(y(k) - \hat{y}(k))^2}{\sum_{k=1}^{N}(y(k) - \bar{y})^2} \tag{34}$$

with the sample mean $\bar{y}$, are employed. To asses the generalization capabilities of the models l-fold cross validation (CV) is used with $l = 10$. In l-fold cross validation the data set is split into l subsets $\{T_1, \ldots, T_l\}$ of equal size. In each iteration $i \in 1, \ldots, l$ the model is trained with $l - 1$ subsets $\{T_1, \ldots, T_l\} \backslash \{T_i\}$ and validated on the remaining subset $\{T_i\}$. The l results are averaged afterwards. The 10-fold cross validated variants of the metrics (33) and (34) are denoted as RMSE_CV and R^2_CV, respectively.

5.3 Results of TS-BE modeling

In Table 1 the model accuracies for the residual stress σ_t are summarized. The identification was conducted with the TS models detailed in section 3.1. The number of submodels $c = 3$, the fuzziness parameter $\nu = 1.3$, and the partition centers $\boldsymbol{\theta}_\text{MF}$:

$$\boldsymbol{v}_1 = [0.586, 0.787, -0.456, -0.107]^T$$
$$\boldsymbol{v}_2 = [0.312, 1.123, -0.576, -0.134]^T$$
$$\boldsymbol{v}_3 = [-2.662, 0.644, 0.035, -0.066]^T$$

were adopted from [23]. The local model parameters $\boldsymbol{\theta}_\text{LM}$ were estimated with the methods from section 3.2.

Table 1: Performance of TS models with BE parameter estimation for σ_t (point predictions with local model parameters $\boldsymbol{\theta}_\text{LM} = \boldsymbol{\theta}_\text{cc}$).

RMSE	RMSE_CV	R^2	R^2_CV
62.097	97.260	0.9292	0.8623

In practice, the size of the resulting feasible parameter set is often determined by few data points, where the actual error is larger than the assumed error bound. This may result, for example, from sensor faults, overoptimistic error bounds or the incapability of the model to explain the system behavior. For the considered uniform error bounds, this can even result in an empty FPS ($\mathbb{S}_\text{FPS} = \emptyset$). As suggested in [22], such critical data points were identified and discarded in a pre-processing step.

For the case study, a simple procedure to discard these data points was applied. The model parameters were estimated based on the point estimation procedure introduced in 3.1. Then 5 % of the training data with the largest residuals were removed. More sophisticated methods for treating points that violate the error bound are possible but are not discussed here for the sake of brevity, see e. g. [16]. The choice of the error bound δ is important as an overoptimistic choice would lead to an empty parameter set while a pessimistic choice would promote a large parameter set. In order to determine an appropriate error bound, δ was iteratively increased within a line-search starting from $\delta = 0.1$ with an increment of 0.01 until $\mathbb{S}_\text{FPS}$ was no longer empty at a value of $\delta = 0.48$.

The result of the BE parameter estimation is the 15-dimensional set of feasible parameters in the representation of its convex hull. To derive a (point-type) parameter vector from $\mathbb{S}_\text{FPS}$ the Chebyshev center $\boldsymbol{\theta}_\text{cc}$, i. e. the center of the maximum volume of the n_θ-dimensional sphere inside the polytope $\mathbb{S}_\text{FPS}$, is used. The Chebyshev center can be computed by solving a linear programming problem [3]. Computation time using TS models implemented in MATLAB and the DD method from MTP3 for vertex enumeration of the full 15-dimensional parameter space was

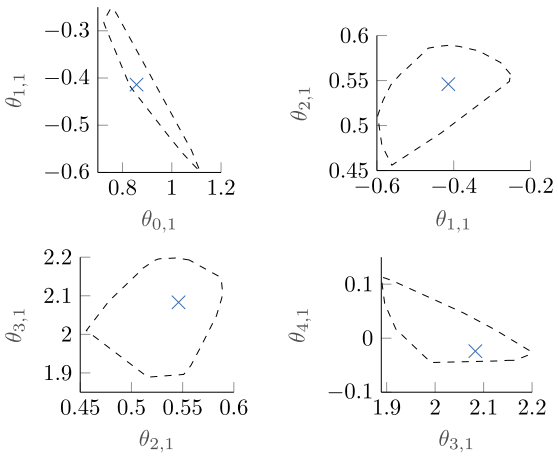


Figure 2: Projections of $\mathbb{S}_{\text{FPS}}$ for the first local model with point estimations (×) using the Chebyshev center.

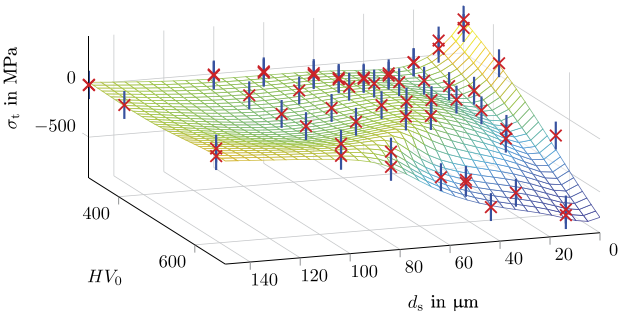


Figure 3: Response surface for σ_t with TS multi-model and BE parameter estimation with measured residual stresses (×) and non-standardized error bounds $\delta = 119$ MPa as blue bars dependent on initial hardnesses HV_0 and distance from surface d_s for constant cutting parameters $f = 0.125$ mm and $v_c = 175$ m/min.

26 s on a workstation with an Intel i5-6500 3.2 GHz and 16 GB RAM.

Figure 2 shows the projections of $\mathbb{S}_{\text{FPS}}$ together with the Chebychev center $\boldsymbol{\theta}_{\text{cc}}$ for one local model as example. Fig. 3 shows the resulting response surface with the error bars for the TS model using the Chebyshev center of $\mathbb{S}_{\text{FPS}}$ as local model parameters $\boldsymbol{\theta}_{\mathbf{LM}} = \boldsymbol{\theta}_{\mathbf{cc}}$, showing a good fit to the measured data.

5.4 Results of GPR modeling

For the GPR modeling, the performance of four different kernel functions were evaluated (SE, M3\2, SE with ARD, M3\2 with ARD) using the MATLAB GPR toolbox. For the quasi Newton optimization algorithm, automatic initial parameter and step size determination was used in MATLAB. The parameter estimation took 29 s using the same hardware as in Section 5.3. In Fig. 4 the response surface with the 95 % confidence interval for the SE and M3\2 kernel are depicted. Table 2 shows that the M3\2 kernel function achieves the best performance on test and training data. It is evident that the ARD improves the results significantly as it eliminates the problem of scale invariance in multivariate GPR.

Table 2: Performance of GPR models for σ_t with different kernel functions.

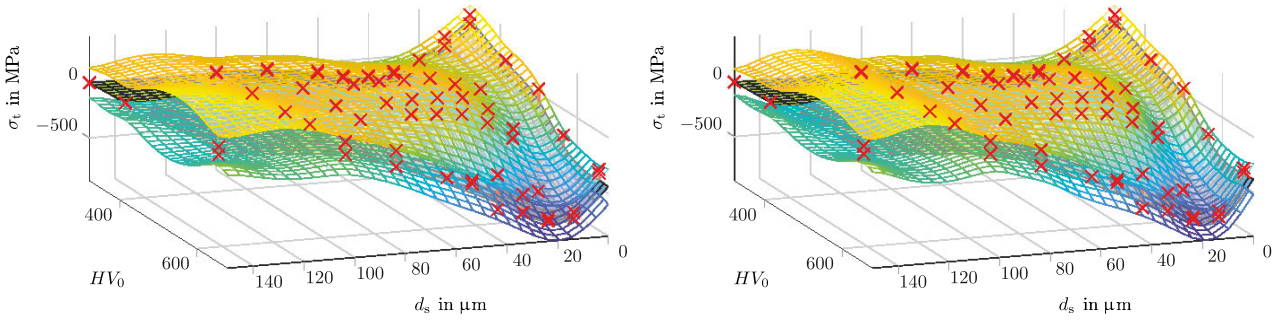
Kernel	RMSE	R^2	RMSE_{CV}	R^2_{CV}
SE	49.403	0.9612	85.3642	0.8872
SE + ARD	39.3029	0.9754	73.8676	0.9120
M3\2	31.0733	0.9848	75.2392	0.9091
M3\2 + ARD	24.5145	0.9905	66.3113	0.9292

Table 3: Estimated noise standard deviations σ_y and values for $1.96 \times \sigma_y$ corresponding to the 95 % confidence interval for different kernel functions for standardized data.

Quantil	SE	SE + ARD	M3\2	M3\2 + ARD
σ_y	0.2530	0.1919	0.2070	0.1823
$1.96 \times \sigma_y$	0.4959	0.3761	0.4057	0.3573

5.5 Discussion

The results indicate that TS models with BE and GPR both well predict the residual stresses. Best results with GPR were achieved with the M3\2 + ARD kernel function. Conducting the Shapiro-Wilk test on the residuals for the GPR models reveals that the normality assumption cannot be accepted at 5 % significance level for all kernel functions. Therefore, the GPR's assumption of normally distributed noise is violated, a problem that often arises in data-driven modeling of real-world processes. The BE approach makes no statistical error assumption, and only requires the error to be bounded. However, single critical data points can cause an empty parameter set as the probability for values outside the interval is zero. In Table 3 the estimated noise standard deviations are given together with the $1.96 \times \sigma_y$ values that correspond to the 95 % confidence interval. The error bound of $\delta = 0.48$, where 5 % of the data with the largest residuals were removed (i. e. 95 % remained in the data set), is similar to the estimated 95 % confidence interval of the GPR. For the case study, the computation time for training is similar for both models: 26 s for the BE parameter estimation for TS and 29 s for the kernel parameter and noise standard deviation estimation for GPR. How-



(a) Response surface (filled mesh) and 95 % confidence envelopes (transparent meshes) with M3\2 kernel function

(b) Response surface (filled mesh) and 95 % confidence envelopes (transparent meshes) with M3\2+ARD kernel function

Figure 4: Response surfaces (filled mesh) for σ_t with 95 %-confidence envelopes (transparent meshes) of GPR models and measured residual stresses (×) dependent on initial hardnesses HV_0 and distance from surface d_s for constant cutting parameters $f = 0.125$ mm and $v_c = 175$ m/min.

ever, memory requirements are much larger for the GPR model as all N data points have to be stored whereas the TS model only requires the $(n + 2) \cdot c$ parameters. Making predictions with the GPR model is computationally more complex, which can be critical in a real time application.

6 Conclusion

Approaches for the prediction of difficult to measure surface properties in machining were introduced. Methods to quantify the data-driven models uncertainty were presented. A nonparametric and a parametric approach both achieved good results. With GPR very good predictive performance on test and training data could be achieved. With the proposed TS model the results were almost as good but had a much more compact and parametric process description. The choice of the appropriate modeling approach is determined by the requirements of the user. The BE approach offers guaranteed error bounds without statistical assumptions while GPR is based on a Gaussian noise assumption and provides a probability distribution. The Gaussian noise assumption is often not met in practical applications. The BE error assumption is much less stringent as it only poses a lower and upper bound. This is the first contribution considering data-driven modeling that quantifies uncertainty for hard turning. The local model approach offers a possibility for using LiP BE parameter estimation for nonlinear regression modeling.

In future work the set-based TS approach will be pursued because of its better suitability for model-based control design compared to GPR models. Advanced methods to treat the critical points will be examined and further work on determining the error bound will be carried out. Furthermore, the model will be extended by surface integrity properties like surface roughness and hardness. Both modeling approaches can easily be transferred to other machining processes.

Funding: The scientific work has been supported by the DFG within the research priority program SPP 2086 (Grant Number: KR 3795/8-1; NI 1327/22-1; ZI 1296/2-1). The authors thank the DFG for this funding and intensive technical support.

References

1. P. J. Arrazola, T. Özel, D. Umbrello, M. Davies, and I. S. Jawahir. Recent advances in modelling of metal machining processes. *CIRP Annals*, 62(2):695–718, 2013.
2. D. Avis and K. Fukuda. A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra. *Discrete & Computational Geometry*, 8(3):295–313, 1992.
3. S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2009.
4. M. H. El-Axir. A method of modeling residual stress distribution in turning for different materials. *International Journal of Machine Tools and Manufacture*, 42(9):1055–1063, 2002.
5. K. Fukuda and A. Prodon. Double description method revisited. In *Combinatorics and Computer Science*, pages 91–111. Springer, 1996.
6. B. Genov. *The convex Hull Problem in Practice: Improving the Runnning Time of the Double Description Method*. PhD thesis, Universität Bremen, 2014.
7. F. Hashimoto, Y. B. Guo, and A. W. Warren. Surface integrity difference between hard turned and ground surfaces and its impact on fatigue life. *CIRP Annals*, 55(1):81–84, 2006.
8. M. Herceg, M. Kvasnica, C. N. Jones, and M. Morari. Multi-Parametric Toolbox 3.0. In *Proc. of the European Control Conference*, pages 502–510. Zürich, Switzerland, 2013.

9. L. Jaulin and E. Walter. Set inversion via interval analysis for nonlinear bounded-error estimation. *Automatica*, 29(4):1053–1064, 1993.
10. D. Kong, Y. Chen, and N. Li. Gaussian process regression for tool wear prediction. *Mechanical Systems and Signal Processing*, 104:556–574, 2018.
11. J. Korbicz and M. Kowal. Neuro-fuzzy networks and their application to fault detection of dynamical systems. *Engineering Applications of Artificial Intelligence*, 20(5):609–617, 2007.
12. A. Kroll. *Computational Intelligence. Probleme, Methoden und technische Anwendungen*. De Gruyter Oldenbourg, 2. edition, 2016.
13. M. Lebsanft, M. Tiffe, A. Zabel, W. Zinn, D. Biermann, and B. Scholtes. Residual stresses in different heat treated workpieces after turning. *Advanced Materials Research*, 996:652–657, 2014.
14. T. S. Motzkin, H. Raiffa, G. L. Thompson, and R. M. Thrall. The double description method. In *Contributions to the Theory of Games (AM-28), Volume II*, pages 51–74. Princeton University Press, 1953.
15. B. Mrugalska. A bounded-error approach to actuator fault diagnosis and remaining useful life prognosis of Takagi-Sugeno fuzzy systems. *ISA Transactions*, 80:257–266, 2018.
16. L. Pronzato and E. Walter. Robustness to outliers of bounded-error estimators, consequences on experiment design. *IFAC Proceedings Volumes*, 24(3):1133–1138, 1991.
17. C. E. Rasmussen and C. K. I. Williams. Gaussian processes for regression. In *Advances in Neural Information Processing Systems 8*, pages 514–520. MIT press, 1996.
18. C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press Ltd, 2005.
19. H. Sasahara. The effect on fatigue life of residual stress and surface hardness resulting from different cutting conditions of 0.45c steel. *International Journal of Machine Tools and Manufacture*, 45(2):131–136, 2005.
20. T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-15(1):116–132, 1985.
21. D. Ulutan and T. Ozel. Machining induced surface integrity in titanium and nickel alloys: A review. *International Journal of Machine Tools and Manufacture*, 51(3):250–280, 2011.
22. E. Walter and H. Piet-Lahanier. Recursive robust minimax estimation for models linear in their parameters. *IFAC Proceedings Volumes*, 25(15):215–220, 1992.
23. F. Wittich, M. Kahl, A. Kroll, W. Zinn, and T. Niendorf. On nonlinear empirical modeling of residual stress profiles in hard turning. In *Proceedings IEEE International Conference on Systems, Man, and Cybernetics*, 2019.
24. G. Zhang, J. Li, Y. Chen, Y. Huang, X. Shao, and M. Li. Prediction of surface roughness in end face milling based on Gaussian process regression and cause analysis considering tool vibration. *The International Journal of Advanced Manufacturing Technology*, 75(9-12):1357–1370, 2014.
25. J. Zhang, S. Liang, G. Zhang, and D. Yen. Modeling of residual stress profile in finish hard turning. *Materials and Manufacturing Processes*, 21(1):39–45, 2006.

Bionotes

M. Sc. Felix Wittich
Department of Measurement and Control, Faculty of Mechanical Engineering, University of Kassel, 34125 Kassel, Germany
felix.wittich@mrt.uni-kassel.de

Felix Wittich is a research associate at the Department of Measurement and Control at the University of Kassel. His research interests include system identification, data-driven modeling, and machine learning.

B. Sc. Lars Kistner
Department of Measurement and Control, Faculty of Mechanical Engineering, University of Kassel, 34125 Kassel, Germany
lars.kistner@mrt.uni-kassel.de

Lars Kistner is a M. Sc. student at the Department of Measurement and Control Engineering at the University of Kassel. His research interests include nonlinear system identification, machine learning, as well as of automation and robotics.

Univ.-Prof. Dr.-Ing. Andreas Kroll
Department of Measurement and Control, Faculty of Mechanical Engineering, University of Kassel, 34125 Kassel, Germany
andreas.kroll@mrt.uni-kassel.de

Andreas Kroll is head of the Department of Measurement and Control Engineering at the University of Kassel. His research interests include nonlinear system identification and control, computational intelligence, complex systems, telemetry and sensor data fusion.

Prof. Dr.-Ing. Christopher Schott
Institute of Materials Engineering, Metallic Materials, University of Kassel, 34125 Kassel, Germany
c.schott@uni-kassel.de

Christopher Schott is a research assistant at the Institute of Materials Engineering at the University of Kassel. His research interests include surface properties, especially residual stresses after turning as well as after straightening operations.

Dipl.-Ing. Thomas Niendorf
Institute of Materials Engineering, Metallic Materials, University of Kassel, 34125 Kassel, Germany
niendorf@uni-kassel.de

Thomas Niendorf is the head of the Metallic Materials group at the Institute for Materials Engineering at the University of Kassel. His research interests are in the interrelationships of process, microstructure, mechanical properties and reliability of metallic materials. Analysis of residual stresses, microstructure evolution and fatigue performance are key aspects of research projects conducted. Materials in focus are steels, aluminum alloys, high-temperature materials, shape memory alloys as well as hybrid materials.