Article

Gaussian Process Surrogates for Modeling Uncertainties in a Use Case of Forging Superalloys

Johannes G. Hoffer [1,*] , Bernhard C. Geiger [1] and Roman Kern [2,*]

1   Know-Center GmbH, Research Center for Data-Driven Business & Big Data Analytics, Inffeldgasse 13, 8010 Graz, Austria; geiger@ieee.org
2   Institute of Interactive Systems and Data Science, Graz University of Technology, Inffeldgasse 16c, 8010 Graz, Austria
*   Correspondence: jhoffer@know-center.at (J.G.H.); rkern@tugraz.at (R.K.)

Abstract: The avoidance of scrap and the adherence to tolerances is an important goal in manufacturing. This requires a good engineering understanding of the underlying process. To achieve this, real physical experiments can be conducted. However, they are expensive in time and resources, and can slow down production. A promising way to overcome these drawbacks is process exploration through simulation, where the finite element method (FEM) is a well-established and robust simulation method. While FEM simulation can provide high-resolution results, it requires extensive computing resources to do so. In addition, the simulation design often depends on unknown process properties. To circumvent these drawbacks, we present a Gaussian Process surrogate model approach that accounts for real physical manufacturing process uncertainties and acts as a substitute for expensive FEM simulation, resulting in a fast and robust method that adequately depicts reality. We demonstrate that active learning can be easily applied with our surrogate model to improve computational resources. On top of that, we present a novel optimization method that treats aleatoric and epistemic uncertainties separately, allowing for greater flexibility in solving inverse problems. We evaluate our model using a typical manufacturing use case, the preforming of an Inconel 625 superalloy billet on a forging press.

Keywords: GP regression; FEM; surrogate modeling; multi-objective optimization; hot metal forming; Inconel 625

1. Introduction

Conducting experiments to better understand manufacturing processes is crucial, with real physical experiments being considered the gold standard. However, conducting real physical experiments for each new experimental setting is impractical because of expensive materials, production stoppages and labor hours for monitoring and evaluation. One good alternative is conducting experiments via simulation, where numerical methods–such as Finite Element Method (FEM)–present a well-observed method in the field of structural analysis. However, solving complex problems with FEM is time-consuming and computationally expensive. In order to reduce the computational effort, surrogate modeling may offer a promising solution [1]. Surrogate models are trained in a supervised manner and are designed to learn the function mapping between inputs and outputs. With a sufficient amount of training data with respect to the observed use case, a customized surrogate model is able to substitute for a FEM simulation up to a certain accuracy. When only specific dimensions with a controlled reduction in accuracy of a simulation result are desired, reduced-order surrogate modeling is an already known technique. Thus, reduced-order surrogate modeling aims to substitute the high-resolution simulation domain with some carefully selected dimensions of importance, e.g., selected displacement measures of a deformed part can be predicted by a reduced-order surrogate modeling with low com-

putational effort, instead of performing a computationally intensive FEM simulation that predicts the displacement of each node representing the deformed part.

Meanwhile, Gaussian process regression (GP) has been successfully used as a surrogate model in the past. In literature, GP regression is also called "kriging" after the statistician and mining engineer Danie G. Krige [2]. However, for consistency, we use only the term GP regression or plain GP in this paper. Regarding GP regression, one of the biggest advantages is that it predicts a distribution (described by mean and standard deviation) rather than just a point estimate. The predicted standard deviation can be seen as a quality criterion related to the corresponding predicted mean value. In the following, we will refer to that standard deviation of a prediction as epistemic uncertainty, i.e., how certain the model is with respect to its prediction. Considering real manufacturing processes, another source of uncertainty can be observed with regard to the lack of complete control over all influence parameters. These deviations occurring during repeated process iterations under the same conditions are referred to as aleatoric uncertainty.

Recapitulating, we want to shed light on two types of uncertainties in surrogate modeling: (1) epistemic uncertainty referring to the lack of knowledge in respect to a simulation model and can be minimized by adding additional sources of information (with respect to machine learning models, it is mainly increasing the number of training instances at new locations in the feature space) and (2) aleatoric uncertainty referring to deviations of an observed manufacturing process itself, i.e., aleatoric uncertainty cannot be minimized even if more data is generated. Since epistemic and aleatoric uncertainties describe different properties, it seems natural to treat them separately when making predictions or optimization. However, it should be mentioned that in certain circumstances it may be useful to consider uncertainty as a whole rather than dividing it into aleatoric and epistemic uncertainty. In such cases, heteroskedastic GP regression represents a common approach for optimization with surrogate models [3–5]. In our problem definition, especially in solving inverse problems, we argue that the distinction of epistemic and aleatoric uncertainty shows clear advantages.

There is a wealth of literature on surrogate models, reduced-order surrogate models, and optimization with GP regression. We present in the following the main related works to our research field organized in (1) GP regression and FEM simulations, (2) GP regression trained with pure sensor data and (3) optimization with GP regression.

In the work of Roberts et al. [6], they predict damage development in forged brake discs reinforced with Al-SiC particles using damage maps. In addition to Multilayer Perceptron (MLP), Roberts et al. [6] utilize GP regression as a surrogate model. Loghin and Ismonov [7] predict the stress intensity factors using GP regression trained with FEM results of a classical bolt-nut assembly use case. Ming et al. [8] model an electrical discharge machining process with GP regression. Su et al. [9] utilize GP regression as a surrogate in a structural reliability analysis of a large suspension bridge. In the work of Guo and Hesthaven [10], GP regression is used as a reduced-order model for nonlinear structural analysis in a 1D and 3D use case, where data generation was performed with active learning. Hu et al. [11] use GP regression to estimate residual stresses field of machined parts from two-dimensional numerical simulations. Yue et al. [12] propose two active learning approaches using GP regression for a composite fuselage use case. In the work of Ortali et al. [13] GP regression is used as a reduced-order surrogate model for fluid dynamics use cases. Venkatraman et al. [14] use GP regression as a surrogate model of texture in micro-springs. GP regression can also be used on data with multiple fidelity levels, where Lee et al. [15] investigate GP regression surrogate modeling with uncertain material properties of soft tissues and multi-fidelity data. Brevault et al. [16] provide an overview of multi-fidelity GP regression techniques in the field of aerospace systems. GP regression can also be extended by methods that stack them or use them in a tree model. Civera et al. [17] predict imperfections in pultruded glass fiber reinforced polymers with a treed method of GP regression trained with experimental data and FEM simulation results. Abdelfatah et al. [18] propose a stacked GP regression to integrate different datasets and propagate uncertainties through the stacked model.

GP regression can also be used for calibrating simulations, where Mao et al. [19] use GP regression as a surrogate model in a Bayesian model updating method to calibrate FEM simulation of a long-span suspension bridge.

In addition to the use of FEM data, GP regression also finds application in the use of pure sensor data, which we will discuss in the following. Tapia et al. [20] use a GP regression based surrogate model of a laser powder-bed fusion process to predict melt pool depth. Yu et al. [21] utilize–besides other thriving methods–a GP regression to model the relationship between geological variables and the broken rock zone thickness. Lee [22] uses GP regression trained with experimental data to optimize wire arc additive manufacturing process deposition parameters. Saul et al. [23] propose chained GP regression models based on non-linear latent function combination. Binois et al. [24] provide a heteroskedastic GP regression approach and results of two use cases, namely manufacturing and management of epidemics.

In the course of function maximization with GP regression surrogate models, Dai Nguyen et al. [25] propose a robust optimization approach based on Upper Confidence Bound (UCB) Bayesian Optimization (BO). In another field of optimization, namely solving inverse problems, there is related work found where BO with generalized chi-squared distribution is researched by Huang et al. [26], and Uhrenholt and Jensen [27], where besides standard GP regression Uhrenholt and Jensen [27] utilized warped GP regression from the work of Snelson et al. [28]. An extension of the standard BO can be found in the work of Plock et al. [29], where they combine BO with the Levenberg-Marquardt method. While in maximization and minimization problems aleatoric and epistemic uncertainties can often be treated in the same way, in most cases robust results can be obtained by distinguishing between these two sources of uncertainty [30]. We refer to robust results when mean predictions are associated with low aleatoric uncertainty.

There is already considerable related work in reduced-order surrogate modeling and optimization using GP regression surrogates. However, to the best of our knowledge, we could not identify related work for solving inverse problems in which aleatoric and epistemic uncertainties are treated differently. Optimization approaches for solving inverse problems usually use only epistemic uncertainty. When epistemic and aleatoric uncertainties are taken into account, they are often simply combined, resulting in the potential loss of important information.

To sum up, we identify the following drawbacks:

- Related work shows that mainly epistemic uncertainty is used for prediction or optimization with GP regression.
- In research using aleatoric and epistemic uncertainties, they are not considered separately when solving inverse problems.

As a response, we present the following main contributions of our research to tackle the identified drawbacks of related work:

1. We present a GP based surrogate that models (a) the mean result, (b) the aleatoric and (c) the epistemic uncertainty of a manufacturing process outcome.
2. We utilize aleatoric and epistemic uncertainties in solving inverse problems for robust optimization results.

With the proposed surrogate model and novel multi-objective optimization strategy, we pave the way for surrogate modeling and inverse problem-solving for practical applications that make use of explicit modeling of sources of uncertainties. Our findings are validated on a typical hot metal forming manufacturing process: preforming an Inconel 625 superalloy billet on a forging press.

This paper is structured as follows. In Section 2, we present the proposed surrogate model, providing an introduction to GP regression in Section 2.1 and describe the GP based parts of our surrogate model in Sections 2.2 and 2.3. The data generation of aleatoric uncertainty for our surrogate model approach is presented in Section 2.4. Section 3 deals with optimization, where we outline active learning in Section 3.1 and solving inverse

problems in Section 3.2. In Section 4 we present the studied use case, preforming an Inconel 625 superalloy billet on a forging press, where we give insights on the design of the forging aggregate characteristics in Section 4.1 and all information regarding the corresponding FEM simulation in Section 4.2. Section 5 shows the results, which are discussed in Section 6. In Section 7, we present the conclusion of our work and an outlook for the future.

2. GP based Surrogate Model

In this section, we first introduce briefly the general idea behind our surrogate modeling approach. We familiarize in Section 2.1 the reader with the general functionality of GP regression to provide an appropriate foundation for the content that follows. In Sections 2.2 and 2.3 we provide more detailed descriptions of each individual GP of our surrogate model. After describing our surrogate model, we move on to uncertainty propagation analysis with FEM simulation in Section 2.4, where we present the procedure for obtaining aleatoric uncertainties.

GP regression is already well researched for surrogate modeling, replacing expensive target labellers (e.g., numerical simulations, expensive manually labelling, conducting real physical experiments, etc.). One reason is their ability to work with low-dimensional data. Another big advantage of using GP regression is that predictions are made in a probabilistic way, i.e., a prediction is represented by a posterior distribution. Thus, a prediction of GP regression is described by a mean and a covariance. The covariance of a prediction can be used as a metric of prediction confidence, i.e., epistemic uncertainty. We specify that outputs of GP regression describe a distribution with mean m and epistemic uncertainty σ.

The proposed surrogate model consists of two individual GPs and takes manufacturing process-specific parameters x_m, part-specific parameters x_p and aleatoric process uncertainty $\bar{\Sigma}_{al}(Z)$ as input and predicts the mean manufacturing result μ and aleatoric uncertainty σ_{al} of the manufacturing result, see Figures 1b, 2 and 3. A similar simulation approach using FEM is shown in Figure 1a. We define Z as a parameter that describes the manufacturing process characteristics, e.g., velocity profile of a forming tool. Our model assumes that $\bar{\Sigma}_{al}(Z)$ can be efficiently obtained for every x_m. This assumption is justified in our running example, where we focus on the first of two directly successive forging strokes. That means that measurements of the manufacturing process are available (i.e., velocity profile of the forging tool), but measurements in respect to the forged part are not possible due to the short time span between the first and second stroke.
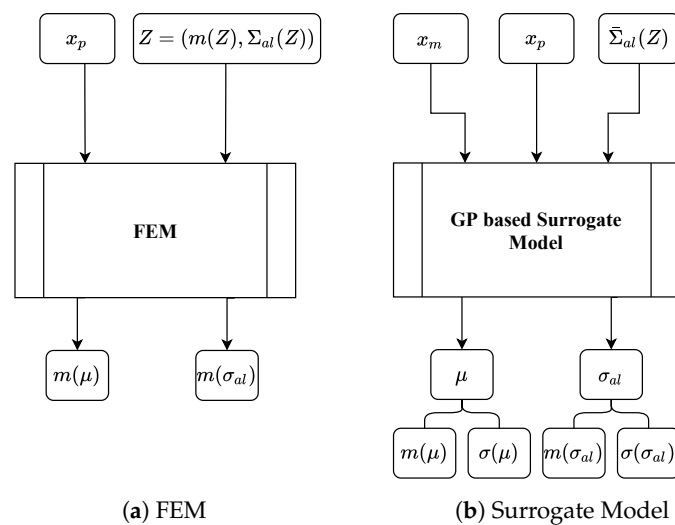

(**a**) FEM   (**b**) Surrogate Model

Figure 1. Simulation of Manufacturing Processes with Uncertainties: (**a**) FEM simulation scenario and (**b**) GP based surrogate model with manufacturing process-specific parameters x_m, part-specific parameters x_p and distribution Z that describes a manufacturing process-specific characteristic by

mean $m(Z)$ and aleatoric manufacturing process uncertainty $\Sigma_{al}(Z)$ where $\bar{\Sigma}_{al}(Z)$ is an aggregated form of $\Sigma_{al}(Z)$. Outputs are the mean of the manufacturing process result $m(\mu)$ and mean of the aleatoric uncertainty $m(\sigma_{al})$, each with corresponding epistemic uncertainties $\sigma(\mu)$ and $\sigma(\sigma_{al})$ in the GP based surrogate model.
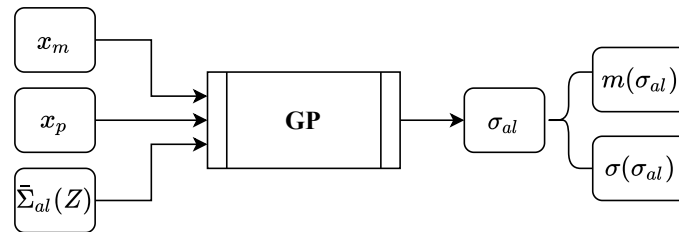

Figure 2. GP takes manufacturing process parameters x_m, part specific parameters x_p and aleatoric manufacturing process uncertainty $\bar{\Sigma}_{al}(Z)$ as input and predicts the mean $m(\sigma_{al})$ and epistemic uncertainty $\sigma(\sigma_{al})$ of the aleatoric uncertainty of the manufacturing process result.
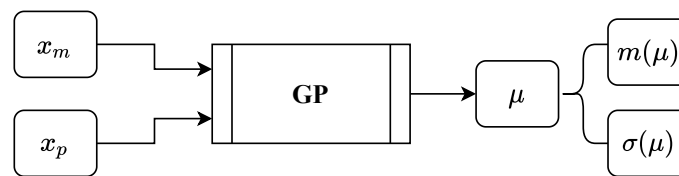

Figure 3. GP takes manufacturing process-specific parameters x_m and part specific parameters x_p as input and predicts the mean $m(\mu)$ and epistemic uncertainty $\sigma(\mu)$ of the manufacturing process result.

2.1. Gaussian Process

A GP is a generalization of the Gaussian distribution. The Gaussian distribution describes random variables or random vectors, while a GP describes functions $f(x)$. In general, a GP is completely specified by its mean function $m(x)$ and covariance function $k(x, x')$, also called kernel. If the function $f(x)$ under consideration is modeled by a GP, we have

$$\mathbb{E}[f(x)] = m(x) \tag{1}$$

$$\mathbb{E}[(f(x) - m(x))(f(x') - m(x'))] = k(x, x') \tag{2}$$

for all x and x'. Where x refers to training and x' to test data. Thus, we can define the GP by

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')). \tag{3}$$

We use the following notation for explanatory purposes only in this section. Matrix $D_{train} = (X, Y)$ contains the training data with input data matrix $X = (x_1, \ldots, x_n)$ and output data matrix $Y = (y_1, \ldots, y_n)$, and test data matrix $D_{test} = (X', Y')$ contains the test data with $X' = (x'_{n+1}, \ldots, x'_{n+m})$ as input and $Y' = (y'_{n+1}, \ldots, y'_{n+m})$ as output. We define that they are jointly Gaussian and have zero mean with consideration of the prior distribution, further, we assume an additive independent identically distributed Gaussian noise with variance σ_n^2 and identity matrix I for noisy observations:

$$\begin{bmatrix} Y \\ Y' \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} k(X, X) + \sigma_n^2 I & k(X, X') \\ k(X', X) & k(X', X') \end{bmatrix}\right) \tag{4}$$

The GP predicts the function values Y' at positions X' in a probabilistic way, where, the posterior distribution can be fully described by the mean and the covariance.

$$\begin{aligned} Y' | X', X, Y \sim \ & \mathcal{N}(k(X', X)[k(X, X) + \sigma_n^2 I]^{-1} Y, \\ & k(X', X') - k(X', X)[k(X, X) + \sigma_n^2 I]^{-1} k(X, X')) \end{aligned} \tag{5}$$

Resulting in mean

$$m(Y') = \mathbb{E}[Y'|X, Y, X'] = k(X', X)[k(X, X) + \sigma_n^2 I]^{-1} Y \tag{6}$$

covariance

$$COV(Y') = k(X', X') - k(X', X)[k(X, X) + \sigma_n^2 I]^{-1} k(X, X') \tag{7}$$

and epistemic standard deviation σ

$$\sigma(Y') = \sqrt{\mathrm{diag}(COV(Y'))} \tag{8}$$

where the diagonal of the covariance matrix COV is extracted as a vector and the square root is calculated for each element to determine the epistemic standard deviation σ. It can be observed that the selection or design of the covariance function is the main ingredient when using GP regression. In the following, we describe the two covariance functions we use in our approach: the popular Radial Basis Function (RBF) (also called squared exponential covariance function)

$$k_{\mathrm{RBF}}(x, x') = \exp\left(\frac{||x - x'||^2}{l^2}\right) \tag{9}$$

with characteristic length-scale parameter l and $|| \cdot ||$ denoting the Euclidean distance and the Matérn covariance function

$$k_{\mathrm{Matérn}}(x, x') = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\sqrt{2\nu}}{l}||x - x'||\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}}{l}||x - x'||\right) \tag{10}$$

with gamma function Γ, modified Bessel function K_ν and parameter ν that controls the smoothness of the resulting function. For more information on GP regression and covariance functions, we refer the reader to the book of Williams and Rasmussen [31].

2.2. Aleatoric Uncertainty GP

A GP is used to predict a manufacturing process related aleatoric uncertainty $\sigma_{al} = \sigma_{al}(x_m, x_p, \bar{\Sigma}_{al}(Z))$ of the manufactured part. Aleatoric uncertainty data are generated by uncertainty propagation analysis with FEM simulation. The inputs are the setting parameters from a real physical manufacturing process x_m, properties of the part to be manufactured x_p and aleatoric manufacturing process uncertainty $\bar{\Sigma}_{al}(Z)$ obtained from, e.g., sensor data of the real physical manufacturing process, see Figure 2. Here, Z describes a characteristic of the manufacturing process, e.g., the velocity profile of a forming tool. The output σ_{al} is predicted by a GP regression, such that $\sigma_{al} \sim \mathcal{GP}(m(x_m, x_p, \bar{\Sigma}_{al}(Z)), k((x_m, x_p, \bar{\Sigma}_{al}(Z)), (x_m, x_p, \bar{\Sigma}_{al}(Z))'))$ with mean $m(x_m, x_p, \bar{\Sigma}_{al}(Z))$ and covariance function $k((x_m, x_p, \bar{\Sigma}_{al}(Z)), (x_m, x_p, \bar{\Sigma}_{al}(Z))')$.

Of course, a wide variety of manufacturing process characteristics can be implemented, e.g., rolling speeds, cutting forces, heating times etc. As a running example, we choose as a manufacturing process hot metal forming on a friction screwpress, where x_m contains different input features which control the forging aggregate (clutch pressure between flywheels and rotation speed of the electric motor), x_p describes the part to be forged by different dimensions and part temperature and Z is a resulting velocity profile of the forging tool over time for a given input x_m, where $\bar{\Sigma}_{al}(Z)$ represents aggregated aleatoric deviations in respect to forging velocity. σ_{al} then describes the deviations of the final forged part, i.e., deviations from important final part geometries. All relevant details of our running example can be found in Section 4.

2.3. Mean Result GP

Besides the GP that predicts the aleatoric uncertainty of a manufactured part, a second GP is used to predict the mean result μ of the manufactured part. The inputs for the second GP are the setting parameters from the real physical manufacturing process x_m and properties of the to be manufactured part x_p. The output μ is predicted by the GP regression, such that $\mu \sim \mathcal{GP}(m(x_m, x_p), k((x_m, x_p), (x_m, x_p)'))$. In respect to our running example, μ describes the final forged part by important final part geometries.

2.4. Uncertainty Propagation Analysis

In uncertainty propagation analysis, the effect of uncertainties related to an input on uncertainties of the corresponding output is investigated. In our case, $\Sigma_{al}(Z)$ refers to the aleatoric deviations of a manufacturing process characteristics (i.e., deviations in velocity profile data) due to different input settings. We refer to uncertainties with respect to a manufacturing process output obtained by uncertainty propagation analysis as aleatoric uncertainty σ_{al}.

We vary input values $x^{(j)} = (x_m^{(j)}, x_p^{(j)})$ with $j \in \{1, \dots, N\}$ where N is the number of different input setting scenarios. For each case of process-specific input parameters $x_m^{(j)}$, we obtain a process-specific characteristic $Z^{(j)}$ that is a distribution with mean $m(Z^{(j)})$ and standard deviation $\Sigma_{al}(Z^{(j)})$. Such distributions occur because, with identical input parameters, process characteristics in reality can show deviations when repeated. We simulate that behavior with a separate GP, thus, a random variable $Z^{(j)}$ is assumed to be Normally distributed, such that $Z^{(j)} = \mathcal{N}(m(Z^{(j)}), \Sigma_{al}(Z^{(j)}))$. From the posterior, we randomly draw M predictions $z^{(i)(j)}$ with $i \in \{1, \dots, M\}$ (i.e., different curves characterizing the manufacturing process) and with each $z^{(i)(j)}$ and $x_p^{(j)}$ we execute FEM simulations to obtain targets $y^{(i)(j)}$. We collect the individual targets $y^{(i)(j)}$, such that we obtain for each input setting j a distribution with mean $\mu^{(j)}$ and aleatoric standard deviation (i.e., aleatoric uncertainty) $\sigma_{al}^{(j)}$. With that, we are able to describe each target by its distribution.

Thus, we obtain a dataset $D = \left\{ D^{(1)}, \dots, D^{(N)} \right\}$ where each datapoint $D^{(j)} = (X^{(j)}, Y^{(j)})$ can then be separated into input $X^{(j)} = (x_m^{(j)}, x_p^{(j)}, \bar{\Sigma}_{al}(Z^{(j)}))$ and output $Y^{(j)} = (\mu^{(j)}, \sigma_{al}^{(j)})$. Here $\bar{\Sigma}_{al}(Z^{(j)})$ is an aggregated manufacturing process uncertainty obtained from data. We model each output with a GP regression, thus the outputs are described again by a distribution with mean m and epistemic standard deviation σ (i.e., epistemic uncertainty), such that $\mu^{(j)} = \mathcal{N}(m(\mu^{(j)}), \sigma(\mu^{(j)}))$ and $\sigma_{al}^{(j)} = \mathcal{N}(m(\sigma_{al}^{(j)}), \sigma(\sigma_{al}^{(j)}))$.

3. Active Learning and Solving Inverse Problems

For optimization, we evaluate our surrogate model in two different areas: (1) active learning and (2) solving multi-objective inverse problems. We refer to active learning as a method to find the most informative data points in the feature space for the best overall performance of the surrogate model, i.e., predicting the mean result of a manufacturing process μ and corresponding aleatoric uncertainty of the manufacturing result σ_{al}. When solving multi-objective inverse problems, we try to find inputs where the error between a given target vector and a prediction as well as the aleatoric uncertainty is minimal, leading to robust optimization results.

3.1. Active Learning

Active learning is already well researched in terms of optimal use of resources for parameter optimization of a model, i.e., generating training data, see [12,32–34]. The process of generating training data means obtaining labels Y_{train} for an input X_{train}, such that a dataset $D_{train} = (X_{train}, Y_{train})$ can be used to fit or optimize parameters of a model. Labels Y_{train} are obtained by an oracle, where an oracle can be a domain expert, results of real physical experiments or like in our case results of expensive numerical FEM simulations.

In the following, we present the idea behind the researched optimization approach and highlight the applicability of active learning with our proposed surrogate model.

In active learning, a number of n_{AL} datapoints connected to maximum epistemic uncertainty σ_{ep} are queried from a pool of candidates X_{pool} to build a training dataset $D_{train} = (X_{train}, Y_{train})$ that is used for training a surrogate model. Thus, we select ideal training data, i.e., we use a minimum amount of training data such that the overall epistemic uncertainty in respect to making prediction on X_{pool} is minimized. We define in (11) the active learning query strategy with loss function $\mathcal{L}_{AL} = \mathcal{L}_{AL}(\sigma_{ep}) = \sigma_{ep}(x)$ to select a new query datapoint $d_q^{AL} = (x_q^{AL}, y_q^{AL})$ with input x_q^{AL} and output y_q^{AL}.

$$d_q^{AL} = \underset{x \ in \ X_{pool}}{\operatorname{argmax}} \sigma_{ep}(x) \tag{11}$$

A query datapoint d_q^{AL} is then moved to the training dataset D_{train}, the surrogate model is fitted and the iterative generation of training data starts again. In respect to our proposed surrogate model, we are able to utilize directly the epistemic uncertainty predictions of the two GPs, i.e., $\sigma(\mu)$ and $\sigma(\sigma_{al})$. Thus, we define $\sigma_{ep}(x) = \sigma(\mu(x)) + \sigma(\sigma_{al}(x))$ and select training data by utilizing (11).

3.2. Inverse Problem

In real physical manufacturing processes, it is commonly required that the result of the manufacturing process lies within a given tolerance range. Therefore, the parameters that control the manufacturing process and the properties of the part must be carefully selected. Moreover, the process of finding inputs to obtain a given target can be formulated as an inverse problem, i.e., finding causal factors for a required effect. In our work, we define that a basic solution of an inverse problem is to find an input x_{inv}, minimizing a distance $d = d(y_{inv}, y_{target})$ between prediction y_{inv} and target vector y_{target}. However, such solutions neglect the existence of process variations, i.e., aleatoric uncertainty. With no consideration of aleatoric uncertainty, the found ideal inputs can lead to quite good results regarding mean values but very high deviations, such that no robustness assertions can be made.

Therefore, we present a novel multi-objective optimization approach in (12) based on BO with a modified UCB acquisition function, where we make a clear separation of uncertainties, such that a loss function $\mathcal{L}_{inv}$, dependent of a distance function d, respective aleatoric σ_{al} and epistemic σ_{ep} uncertainties is minimized.

$$x_{inv} = \underset{X_{pool}}{\operatorname{argmin}} \mathcal{L}_{inv}(d, \sigma_{al}, \sigma_{ep}) \tag{12}$$

As a distance function d, we select the absolute error between mean target μ_{target} and mean manufacturing process result $m(\mu)$ as the metric. However, our approach is not limited to a specific distance metric, so any can be used.

$$d = d(\mu_{target}, m(\mu)) = |\mu_{target} - m(\mu)| \tag{13}$$

Utilizing $m(\sigma_{al})$, $\sigma(\sigma_{al})$, $m(\mu)$ and $\sigma(\mu)$ from our proposed surrogate model, we define epistemic uncertainty $\sigma_{ep} = \sigma_{ep}(\sigma(\sigma_{al}), \sigma(\mu)) = \sigma(\sigma_{al}) + \sigma(\mu)$ and construct a loss function $\mathcal{L}_{inv}$ with tuning parameters α and β, where α controls the influence of the aleatoric uncertainty and β controls exploration vs. exploitation, i.e., the influence of the epistemic uncertainty.

$$\mathcal{L}_{inv}(d(\mu_{target}, m(\mu)), m(\sigma_{al}), \sigma_{ep}(\sigma(\sigma_{al}), \sigma(\mu))) = \\ d(\mu_{target}, m(\mu)) + \alpha \cdot m(\sigma_{al}) - \beta \cdot \sigma_{ep}(\sigma(\sigma_{al}), \sigma(\mu)) \tag{14}$$

Thus, with our approach, we find inputs that provide robust outputs close to a given target while keeping aleatoric uncertainty low. As a result, we obtain robust optimization outcomes when solving multi-objective inverse problems with our approach. In the work of

Dai Nguyen et al. [25] we found a similar handling of uncertainty in the observation of the acquisition function, however, the authors only focus on maximizing black box functions, while we present an extension that solves multi-objective inverse problems.

4. Case Study on Forging Superalloys

We evaluate the proposed surrogate model and novel optimization method with a classic use case from the field of hot metal forming, preforming an Inconel 625 superalloy billet on an artificially designed forging press. First, we design the forging press characteristics with a parameterized curve and a GP and second, we design the forming process itself in a FEM simulation environment where we provide all the relevant information so that it is possible for researchers to link directly to our work.

4.1. Forging Aggregate Characteristic

We calculate the mean forming velocity values of an artificially designed forging process on the example of a forging screwpress by a self-designed parameterized curve in (15) that models the die velocity v_{die} in mm/s as a function of the process timestep t in *seconds*, clutch pressure x_1 in *bar* and rotation speed of the electric motor x_2 in rpm, such that $v_{die} = v_{die}(x_1, x_2, t)$. Where, x_1 and x_2 are two process-specific setting parameters, i.e. $x_m = (x_1, x_2)$.

$$v_{die}(x_1, x_2, t) = \kappa_1 \cdot x_1 \cdot x_2 \cdot t^2 - \kappa_2 \cdot x_1 \cdot x_2 \cdot t^3 \tag{15}$$

where $\kappa_1 = \frac{5}{3}$ mm^2/kg and $\kappa_2 = \frac{5}{3}$ mm^2/kgs are constants. We utilize a designed forging press specific GP with data generated by using (15) to model the mean and input dependent deviations in respect to the manufacturing process characteristic Z (i.e., Z represents the velocity profile of the forging die v_{die}). Z is defined by a distribution with mean $m(Z)$ and aleatoric standard deviation $\Sigma_{al}(Z)$. With respect to our use case, the forging press specific GP with output $Z^{(j)}$ is at the very beginning of the uncertainty propagation analysis, see Figure 4. The inputs for the forging press GP are $x_m = (x_1, x_2)$ and time increments $t = \{0, \ldots, T\}$, where T represents the duration of the manufacturing process. The output of the forging press GP is Z, such that $Z \sim \mathcal{GP}(m(x_m, t), k((x_m, t), (x_m, t)'))$ with mean $m(x_m, t)$ and covariance function $k((x_m, t), (x_m, t)')$. Thus, we obtain for each time increment a distribution describing the velocity at time t. The principle GP design for the forging press can be seen in Figure 5. As covariance function, k we found out that an RBF kernel is appropriate.
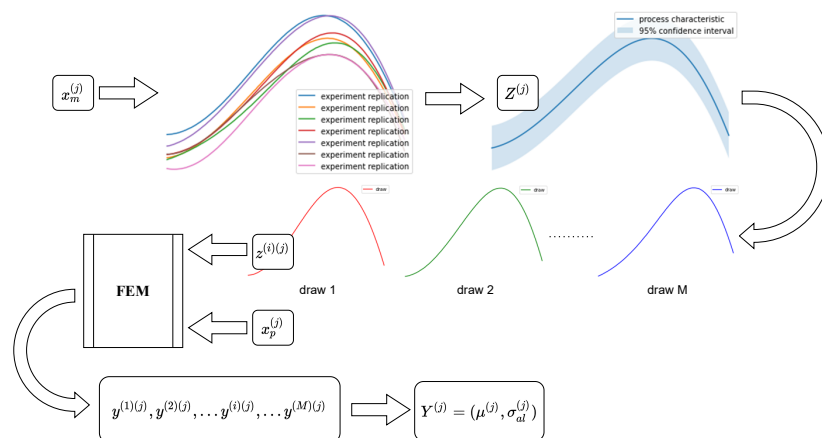

Figure 4. Uncertainty Propagation Analysis: the characteristic $Z^{(j)}$ of the manufacturing process is described by a distribution, since deviations occur when the process is repeated with identical $x_m^{(j)}$. With M draws of $z^{(i)(j)}$ out of the distribution $Z^{(j)}$ as manufacturing process characteristic and to be manufactured part parameters $x_p^{(j)}$, FEM simulations are executed to obtain targets $y^{(i)(j)}$ that describe a distribution $Y^{(j)} = (\mu^{(j)}, \sigma_{al}^{(j)})$ with mean $\mu^{(j)}$ and aleatoric standard deviation, i.e., uncertainty $\sigma_{al}^{(j)}$ for given inputs $x_m^{(j)}$ and $x_p^{(j)}$.
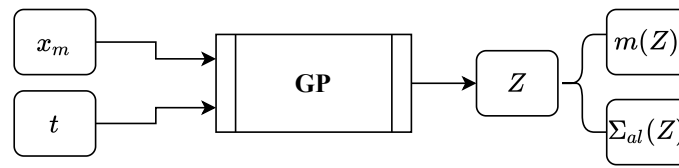
Figure 5. GP takes manufacturing process-specific parameters x_m and manufacturing process time steps t as input and predicts a manufacturing process-specific characteristic (i.e., velocity profile of the forging die) Z with mean $m(Z)$ and uncertainty $\Sigma_{al}(Z)$.

We utilize (15) and different input parameter combinations to generate training data for the forging press GP, see Table 1. In terms of time step size t, we assume that each forging stroke lasts one second, and we model each stroke with a resolution of 100 time steps.

Table 1. Input parameter combinations to generate training data for the forging press GP.

Training Data for Forging Press GP								
x_1	12	12	12	16	16	16	20	20	20
x_2	50	60	70	50	60	70	50	60	70

To obtain different deviations connected to different x_1 and x_2 combinations, we use the underlying inference properties of GP regression and vary inter- and extrapolation tasks in respect to the input values for forging process representation, see Table 2.

Table 2. Input parameter combinations to model forging press characteristics.

Evaluation Data for Forging Press Characteristics															
x_1	10	10	10	10	14	14	14	14	18	18	18	18	22	22	22	22
x_2	45	55	65	75	45	55	65	75	45	55	65	75	45	55	65	75

We define interpolation such that a value is within the training range (i.e., x_1 equals 14 or 18 and x_2 equals 55 or 65) and extrapolation such that a value is out of the training range (i.e., x_1 equals 10 or 22 and x_2 equals 45 or 75).

Exemplary forging press characteristics can be seen in Figure 6, where Figure 6a shows low deviation because x_1 and x_2 are both lie within the range of training data, Figure 6b,c show moderate deviation because one of the process parameters is within and the other is outside the range of the training data and Figure 6d shows the highest deviation because both of the process-parameters lie outside the range of training data. Thus, our forging press GP represents a forging aggregate characteristics with uncertainties dependent on the inputs. In our approach, we intentionally generate deviations depending on input parameters and assume that uncertainty is aleatoric to approximate reality, i.e., we abuse epistemic uncertainty and assume that it is aleatoric. When working with sensor data coming from a real manufacturing process, it is obvious that deviations, i.e., aleatoric process uncertainties Σ_{al}, can be directly measured from data.
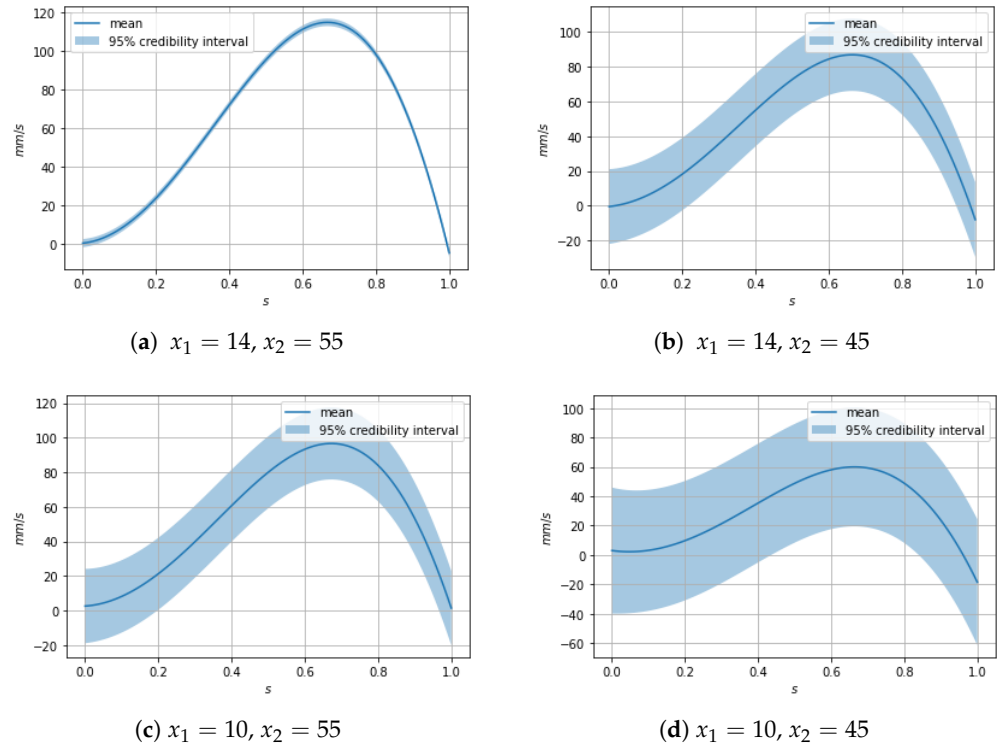
(**a**) $x_1 = 14$, $x_2 = 55$

(**b**) $x_1 = 14$, $x_2 = 45$

(**c**) $x_1 = 10$, $x_2 = 55$

(**d**) $x_1 = 10$, $x_2 = 45$

Figure 6. Exemplary forging press characteristics $Z^{(j)}$ represented by mean and 95% credibility interval of v_{die} over t with (**a**) low deviation, (**b**,**c**) moderate deviation and (**d**) high deviation.

4.2. FEM Simulation

The considered use case, preforming an Inconel 625 superalloy billet on a forging press machine, is observed by utilizing a corresponding FEM simulation. Manufacturing process related FEM inputs $Z^{(j)}$, i.e., different velocity profiles of the upper die over time, are modeled by the forging press GP. Inputs for the forging press GP are x_1, x_2 and t, such that $Z^{(j)(t)} = Z^{(j)(t)}(x_1^{(j)}, x_2^{(j)}, t)$. All 16 possible combinations for manufacturing process related FEM inputs are shown in Table 2. Billet related inputs $x_p^{(j)}$ that are shared with our proposed surrogate model and FEM simulation are diameter, height and temperature, such that $x_p^{(j)} = (d^{(j)}, h^{(j)}, \theta^{(j)})$. One possible billet configuration is shown in Figure 7 and possible billet parameters for different configurations are shown in Table 3. We define the radius of the rounded edges to be constant 10 mm across all configurations.

Table 3. Key parameters for billet configurations $x_p^{(j)}$, values in mm.

	Configuration Data		
Diameter d	220	240	260
Height h	200	210	220
Temperature θ	900	1000	1100

We observe in total 27 different billets. Connecting manufacturing process related combinations with different billets, we obtain 432 combinations, i.e., $j \in \{1, \dots, 432\}$. For evaluation of the uncertainty propagation, we randomly draw $z^{(i)(j)}$ with $i \in \{1, \dots, 20\}$ from each distribution $Z^{(j)}$, i.e., 20 FEM simulations are performed for each input setting. Thus, a total of 8640 FEM simulation results are generated for our experiments. Selected FEM output variables for our surrogate model are the final diameter and height of the preformed billet, such that $y^{(i)(j)} = (d_{final}^{(i)(j)}, h_{final}^{(i)(j)})$ and $Y^{(j)} = (\mu(d_{final}^{(j)}), \mu(h_{final}^{(j)}), \sigma_{al}(d_{final}^{(j)}), \sigma_{al}(h_{final}^{(j)}))$.

In respect to the final diameter d_{final}, we calculate the empiric mean by $\mu(d_{final}^{(j)}) = \frac{1}{20}\sum_{i=1}^{20} d_{final}^{(i)(j)}$ and aleatoric standard deviation by $\sigma_{al}(d_{final}^{(j)})^2 = \frac{1}{20}\sum_{i=1}^{20}(d_{final}^{(i)(j)} - \mu(d_{final}^{(j)}))^2$. The calculations are analogous with respect to h_{final}. Thus, we obtain a dataset with 432 instances described by six input features and four output features. For our running example, input features are clutch pressure, rotation speed, initial billet diameter, initial billet height, initial billet temperature and aggregated manufacturing process uncertainties obtained from data, i.e., the aggregated output of the forging press GP $\bar{\Sigma}_{al}(Z^{(j)}) = \sum_{t=1}^{T}\Sigma_{al}(Z)^{(j)(t)}$. Output features are the mean of the final billet diameter, the mean of the final billet height, the aleatoric uncertainty of the final billet diameter and the aleatoric uncertainty of the final billet height.
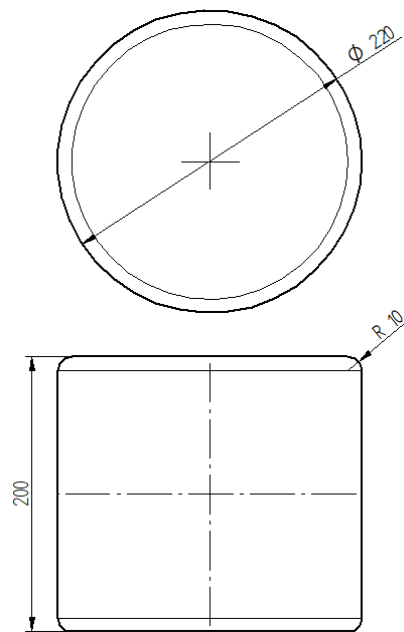

Figure 7. Billet configuration with Diameter $d = 220$ mm, Height $h = 200$ mm and rounded edges with Radius = 10 mm.

The problem is defined as a 2D axisymmetric simulation task to utilize symmetries and make efficient use of computational resources. We utilize isotropic elasto-plastic Inconel 625 material behavior from literature. The Young's modulus is temperature-dependent and the yield stress depends on plastic strain, strain-rate and temperature. We set contact properties to tangential behavior with isotropic directionality and a friction coefficient of 0.3 between the billet and upper and lower forging tool, which means that we assume lubricated hot forging conditions. The lower tool is encastred and the upper tool's boundary conditions are set so that the vertical movement $z^{(i)(j)}$ is drawn from distribution $Z^{(j)}$ and there is no horizontal movement. An exemplary simulation definition can be seen in Figure 8, where (a) shows the initial state of the billet loaded with a randomly drawn screwpress velocity profile $z^{(i)(j)}$ and (b) the end result of the simulation with selected FEM output variables $y^{(i)(j)}$, i.e., the final diameter of 288 mm and the final height of 92.83 mm.

All billets are meshed with an approximate global element size of 7 mm, using 4-node bilinear axisymmetric quadrilateral elements with reduced integration and hourglass control. We obtain our FEM simulation results in the context of general static simulations. Details of the simulation steps are shown in Table 4. Simulation control parameters that are not listed are left at default values.
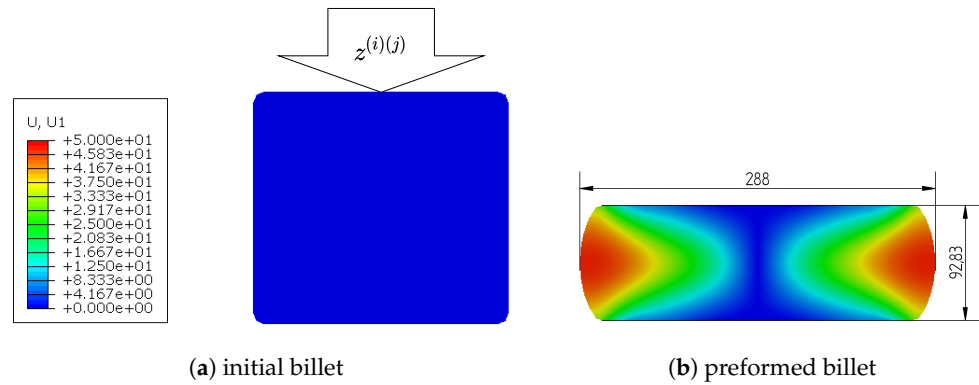
(**a**) initial billet   (**b**) preformed billet

Figure 8. Preforming an Inconel 625 superalloy billet: (**a**) initial billet and randomly drawn velocity profile $z^{(i)(j)}$, (**b**) FEM simulation result with graphical presentation of the horizontal displacement $U, U1$ and selected output variables $y^{(i)(j)}$, i.e., final diameter of 288 mm and final height of 92.83 mm.

Table 4. Abaqus FEM simulation control parameters for our use case.

Abaqus FEM Simulation Settings	
Simulation type	Static, General
Time period	1
Nlgeom	On
Max number of increments	1000
Initial increment size	0.001
Min increment size	1×10^{-5}
Max increment size	1
Equation solver method	Direct
Solution technique	Full Newton

5. Results

5.1. GPs

Before utilizing optimization methods, we evaluate each individual GP, see Table 5. The screwpress GP is trained with data that is generated by using inputs from Table 1 with (15) and tested on data generated by using inputs from Table 2 with (15). As covariance function, k we found out that an RBF kernel is appropriate for this GP. The GPs of our proposed surrogate model are both designed with a Matérn kernel with $\nu = 2.5$ and are independently evaluated by 10-fold cross-validation with inputs from Table 2 and $Z^{(j)}$ obtained from the screwpress GP. Outputs are obtained from FEM simulations, see Section 4.2. In each cross-validation step, we split the respective data randomly such that 10 percent are in the test dataset and the remaining 90 percent are used for model training.

Table 5. Evaluation of individual GPs by average R2-Scores over 10 folds.

Individual GP Evaluations					
	Screwpress	Aleatoric Uncertainty		Mean Result	
	$m(Z)$	$m(\sigma_{al,d_{final}})$	$m(\sigma_{al,h_{final}})$	$m(\mu_{d_{final}})$	$m(\mu_{h_{final}})$
R2-Score	0.9923	0.8146	0.8455	0.9586	0.9555

In addition, we calculate mean Pearson kurtosis

$$kurt_{Pearson} = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{M} \sum_{i=1}^{M} \left(\frac{y^{(i)(j)} - \mu^{(j)}}{\sigma_{al}^{(j)}} \right)^4 \qquad (16)$$

and mean Fisher-Pearson coefficient of skewness

$$skew_{Fisher-Pearson} = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{M} \sum_{i=1}^{M} \left(\frac{y^{(i)(j)} - \mu^{(j)}}{\sigma_{al}^{(j)}} \right)^3 \tag{17}$$

to describe the distribution shapes obtained from uncertainty propagation analysis, see Table 6.

Table 6. Mean values of Pearson kurtosis and Fisher-Pearson coefficient of skewness calculated from uncertainty propagation analysis results.

	Distribution Properties	
	d_{final}	h_{final}
$kurt_{Pearson}$	3.003	2.685
$skew_{Fisher-Pearson}$	0.449	0.015

GP models were implemented with the GPflow library version 2.2.1 and Python 3.8.10. Inferences were run on a machine with 16 GB RAM, 8 CPUs and Intel(R) i7-8565 2.0 GHz processor. We utilized a L-BFGS-B algorithm to train the models. Training our surrogate model on all available data took an average of 1.36 s based on 10 measurements. For one prediction our model needs on average 0.046 s. A FEM simulation lasted on average 149.78 s.

5.2. Active Learning

We evaluate our proposed surrogate model by using active learning and compare it with an approach based on random training data selection. Evaluation is based on 10-fold cross-validation. In each cross-validation step, models are initially trained on two randomly selected datapoints out of the pool dataset containing 432 instances. Evaluation metrics are R2-Score and mean-squared-error (MSE) and are computed on a 20 percent hold-out test set that is randomly generated in each cross-validation step. Results for the mean of reduced-order predictions and corresponding aleatoric uncertainties regarding final diameter and height are shown respectively in Figures 9 and 10, where solid lines depict the mean R2-Score values and shaded areas are obtained by adding and subtracting one standard deviation. Mean values and standard deviations are calculated from the 10 cross-validation results.

5.3. Solving Inverse Problem

We evaluate our proposed multi-objective optimization strategy by solving inverse problems, i.e., we try to find input settings that lead to an output that is as near as possible to an initially defined target vector. In addition to minimize distances between a target vector y_{target} and random mean vector $m(\mu^{(j)})$, we try to achieve results that also keep mean aleatoric uncertainty $m(\sigma_{al}^{(j)})$ low. We utilize a 10-fold cross-validation, where in each cross-validation step the target vector is randomly drawn from the pool dataset and the best found prediction after drawing 50 datapoints out of the pool dataset is used for evaluation. This means that for each method, a dataset of 50 datapoints is generated, and each best prediction is found by evaluating the respective acquisition function on the corresponding generated dataset. We compare our approach with two other baselines, namely:
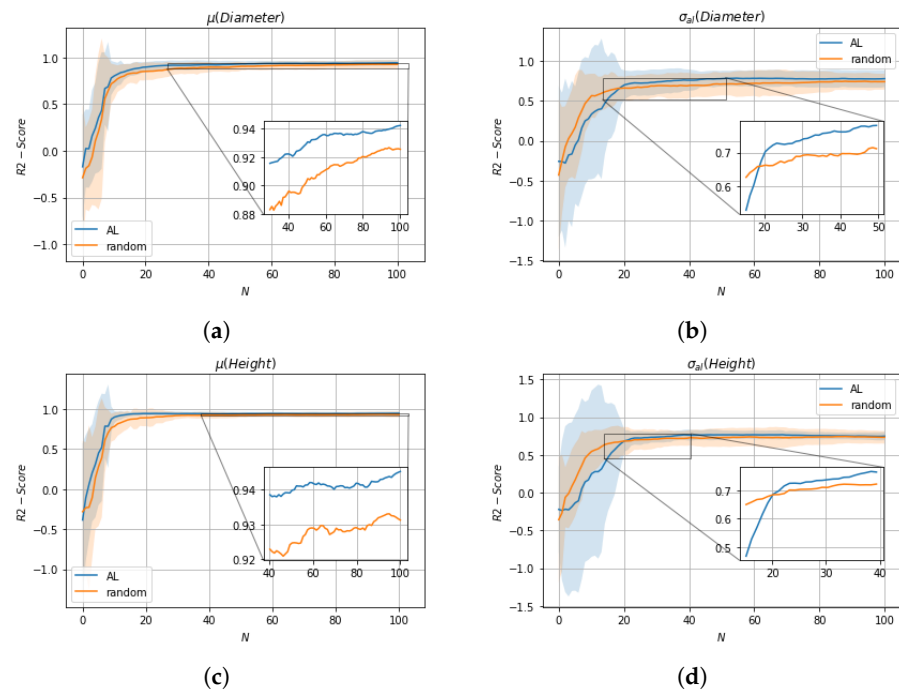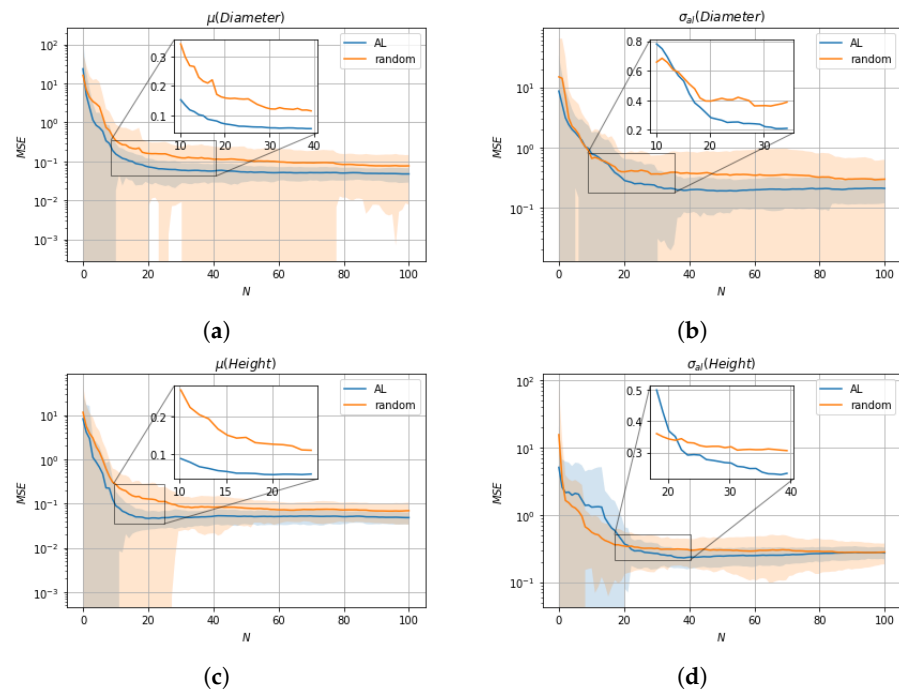
Figure 9. R2-Scores of 10-fold cross-validation over number of drawn training data *N*. In each cross-validation step, models are initially trained on two randomly selected datapoints drawn from the pool dataset. Solid lines depict the mean R2-Score values and shaded areas the upper and lower confidence bounds obtained by adding and subtracting the standard deviations, calculated from the obtained results. (**a**) $m(\mu_{Diameter})$; (**b**) $m(\sigma_{al,Diameter})$; (**c**) $m(\mu_{Height})$; (**d**) $m(\sigma_{al,Height})$.

Figure 10. MSEs of 10-fold cross-validation over number of drawn training data *N*. In each cross-validation step, models are initially trained on two randomly selected datapoints drawn from the pool dataset. Solid lines depict the mean R2-Score values and shaded areas the upper and lower confidence bounds obtained by adding and subtracting the standard deviations, calculated from the 10 obtained results. (**a**) $m(\mu_{Diameter})$; (**b**) $m(\sigma_{al,Diameter})$; (**c**) $m(\mu_{Height})$; (**d**) $m(\sigma_{al,Height})$.

1. Combined (This baseline can be considered as an approximation to the use of heteroskedastic GP in UCB BO.): no distinction of uncertainties in UCB based BO, i.e., simply adding aleatoric and epistemic uncertainty with loss:

$$\mathcal{L}_{combined}(d(\mu_{target}, m(\mu)), m(\sigma_{al}), \sigma_{ep}(\sigma(\sigma_{al}), \sigma(\mu))) =$$
$$d(\mu_{target}, m(\mu)) - [\alpha \cdot m(\sigma_{al}) + \beta \cdot \sigma_{ep}(\sigma(\sigma_{al}), \sigma(\mu)))] \quad (18)$$

2. Epistemic: neglecting aleatoric uncertainty in UCB based BO with loss:

$$\mathcal{L}_{epistemic}(d(\mu_{target}, m(\mu)), m(\sigma_{al}), \sigma_{ep}(\sigma(\sigma_{al}), \sigma(\mu))) =$$
$$d(\mu_{target}, m(\mu)) - \beta \cdot \sigma_{ep}(\sigma(\sigma_{al}), \sigma(\mu)). \quad (19)$$

Figure 11 shows representative plots of optimization results for one random target vector (i.e., one cross-validation step) over 50 draws of x_{inv}, where solid lines depict squared errors and dotted lines show mean aleatoric uncertainty $m(\sigma_{al})$. Figures 12 and 13 show different distributions of optimization results obtained by 10-fold cross-validation in respect to squared errors and mean aleatoric uncertainty $m(\sigma_{al})$. Distributions are visualized by kernel density estimation.
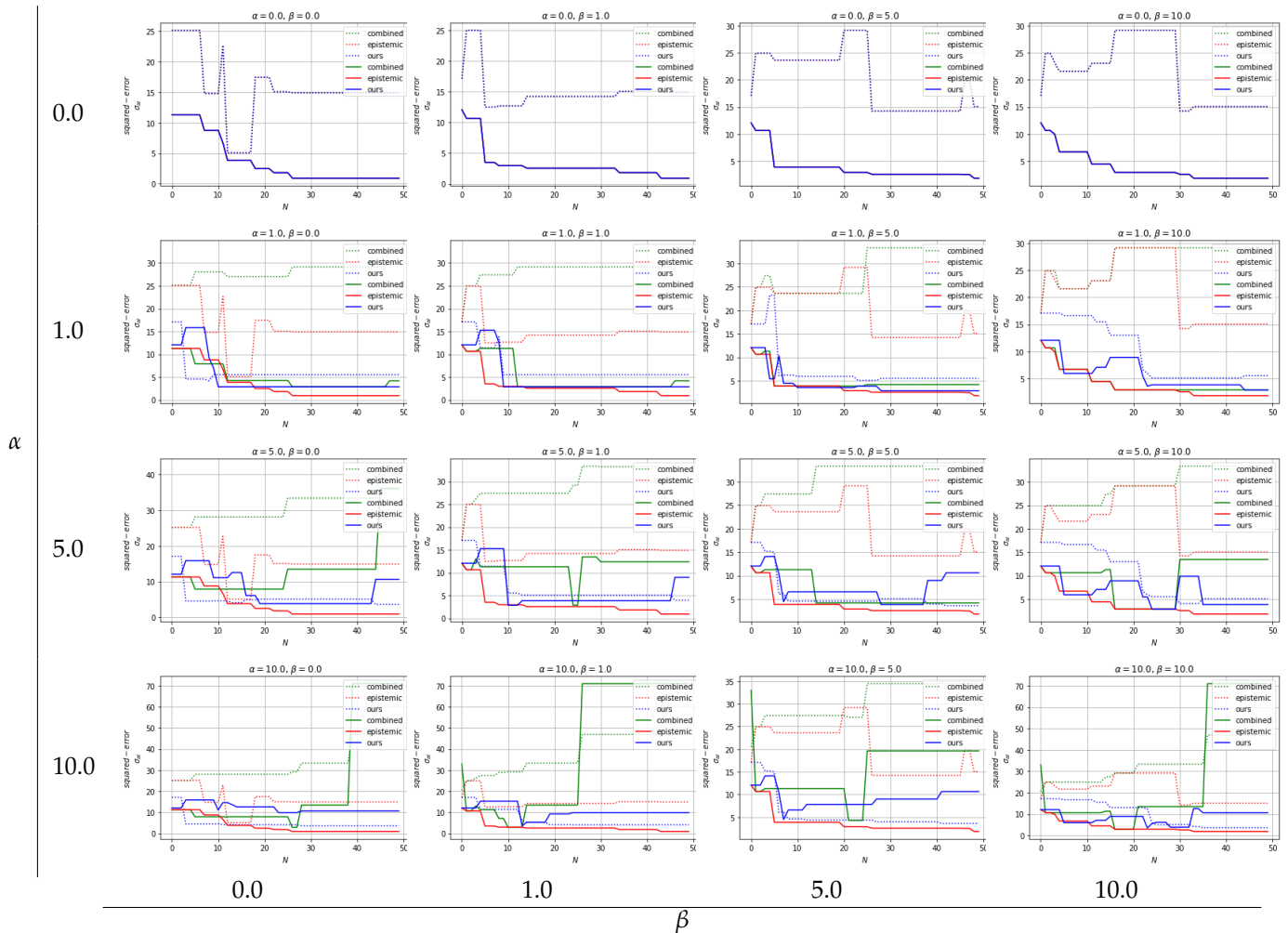

Figure 11. Representative plots of multi-objective optimization results for different hyperparameter settings α and β over number of optimization steps N. Solid lines depict squared error values, and dotted lines represent corresponding mean aleatoric uncertainty $m(\sigma_{al})$. The plots for $\alpha = 0$ show only blue lines, because the results of the different methods are the same and the lines are on top of each other.
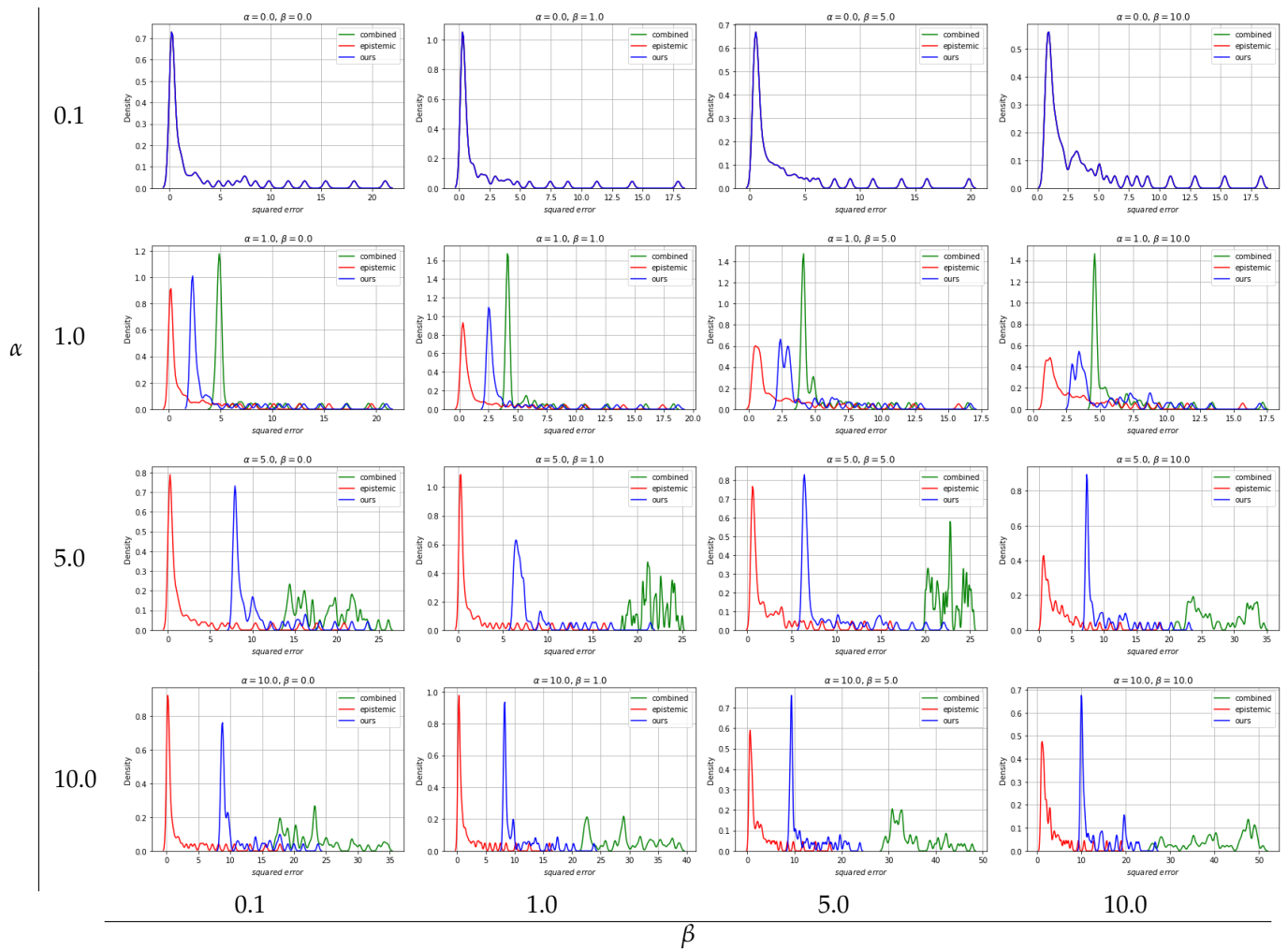
Figure 12. Kernel density estimate plots of squared errors for different hyperparameter settings α and β, distributions are obtained by 10-fold cross-validation, where in each fold a target vector is randomly selected. The plots for $\alpha = 0$ show only blue lines, because the results of the different methods are the same and the lines are on top of each other.
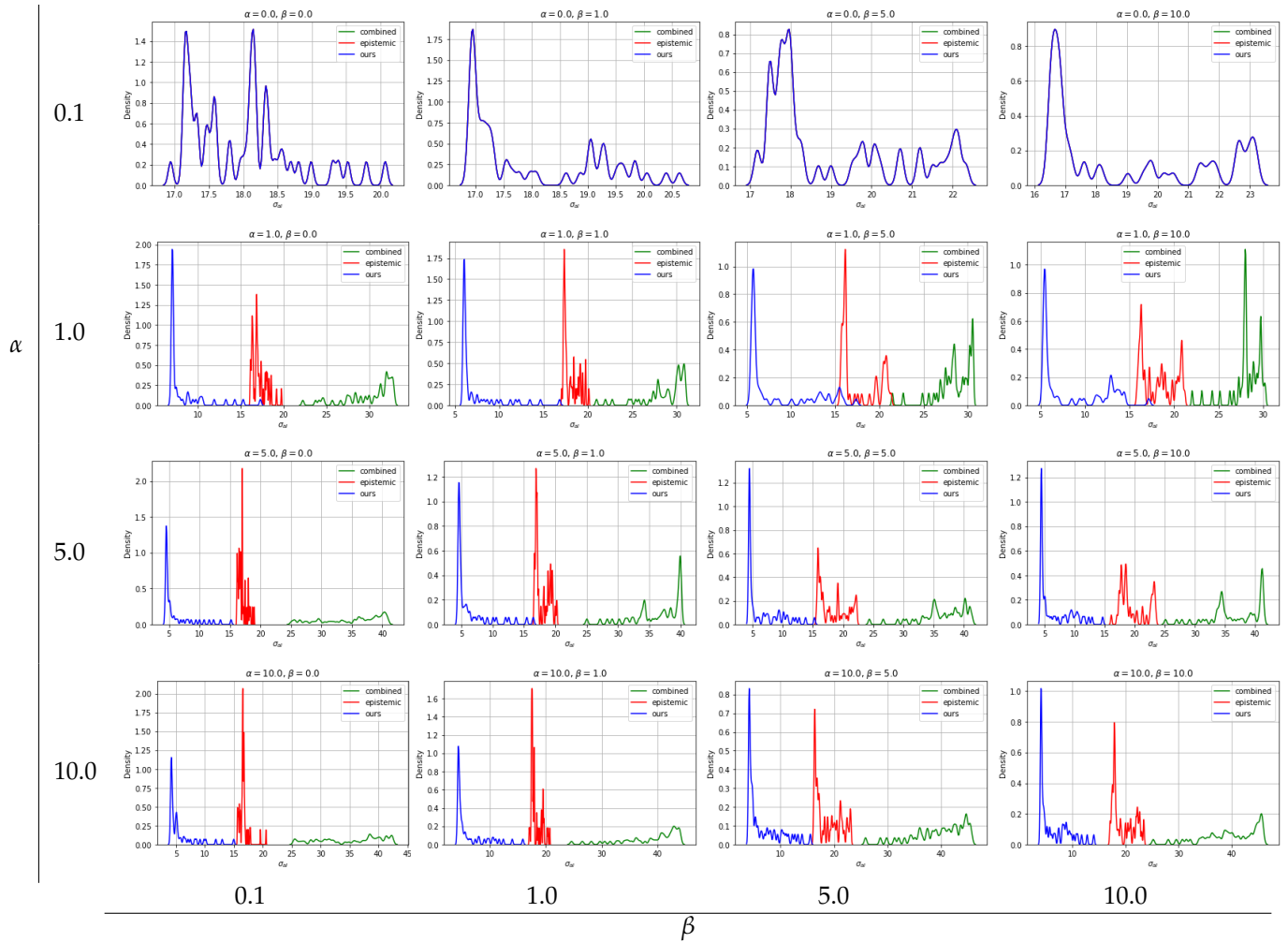
Figure 13. Kernel density estimate plots of mean aleatoric uncertainties $m(\sigma_{al})$ for different hyperparameter settings α and β, distributions are obtained by 10-fold cross-validation, where in each fold a target vector is randomly selected. The plots for $\alpha = 0$ show only blue lines, because the results of the different methods are the same and the lines are on top of each other.

6. Discussion

Evaluation of the individual GPs with 10-fold cross-validation shows promising R2-Scores (lowest: 0.8146, mean: 0.89355, highest: 0.9586), i.e., hyperparameters appear to be appropriate for further evaluations. Observation of generated manufacturing process uncertainties, i.e., $\bar{\Sigma}_{al}(Z^{(j)})$ shows a diverse data landscape, thus, we assume that further uncertainty propagation analysis is meaningful.

We observe the distributions obtained from uncertainty propagation analysis by calculating Pearson kurtosis and Fisher-Pearson coefficient of skewness (A Pearson kurtosis of 3.0 and Fisher-Pearson coefficient of skewness of 0.0 describe a normal distribution). Regarding kurtosis, results shows that distributions are near to Normal distributions, where the distribution of h_{final} is slightly platykurtic ($kurt_{Pearson}(h_{final}) = 2.685 < 3.0$), i.e., it is less peaked than a Normal distribution and the distribution of d_{final} is little leptokurtic ($kurt_{Pearson}(d_{final}) = 3.003 > 3.0$), i.e., the distribution is more peaked compared to a Normal distribution. In terms of skewness, the distribution of d_{final} is more skewed compared to h_{final}, however, both values are less than 0.5 so that approximate symmetry can be assumed.

We evaluate the impact of data selection for model training using two metrics, R2-Score and MSE, with a 10-fold cross-validation comparing active learning with random sample

selection. With respect to mean values μ, active learning shows overall an improvement compared to random sample selection. In terms of aleatoric uncertainties σ_{al}, random sample selection is superior to active learning up to the selection of about 20 samples, but after that active learning shows superior performance compared to random sample selection. The initially worse performance of active learning with respect to σ_{al} is due to a trade-off in the active learning cost function between $\sigma(\mu)$ and $\sigma(\sigma_{al})$ with the influence of $\sigma(\mu)$ dominating. A possible solution for this would be the introduction of appropriate tuning parameters that regulate the influence of the respective epistemic uncertainties $\sigma(\mu)$ and $\sigma(\sigma_{al})$. Moreover, it should be noted that random sample selection shows only better performance at a stage where the tuning of parameters is far from complete, so the better performance is not applicable in practice.

With regard to solving inverse problems, we compare our novel robust UCB based BO multi-objective optimization algorithm with two baselines: (1) combined: no distinction of uncertainties in UCB based BO and (2) epistemic: neglecting aleatoric uncertainty in UCB based BO. We show that over different values of tuning parameters α and β there are clear tendencies of the different approaches. By disabling the influence of aleatoric uncertainty ($\alpha = 0$), all three approaches show the same results as expected: low squared errors and neglected aleatoric uncertainty. For all approaches, slight differences can be seen over different β values while $\alpha = 0$, regulating the exploration vs. exploitation trade-off.

Due to the fundamentals of the epistemic approach, there are no differences in the optimization result when α values are changed for constant β values, see Figure 11. Differences in kernel density estimate plots over varying α values are from random target vector selection. Overall, the epistemic approach yields the best optimization results in terms of squared errors, see Figures 11 and 12, however, as expected, aleatoric uncertainty is ignored and thus high, see Figure 13. The combined approach, where aleatoric and epistemic uncertainties are simply added and handled as quasi-epistemic, shows the overall worst results. At low α values, the squared errors are acceptable, but the aleatoric uncertainty is high due to inappropriate handling of information, see Figures 11–13. To arrive at our approach, once aleatoric uncertainty is considered, i.e., $\alpha > 0.0$ results for the inverse problem show low squared errors and low aleatoric uncertainty which we recognize as robust results. Moreover, by increasing α one can see that our approach leads to results where lowering aleatoric uncertainty σ_{al} is more preferred than lowering squared errors, see Figure 11 $\alpha = 1.0$ and $\alpha = 10.0$. Kernel density estimate plots generated from 10-fold cross-validation results confirm those findings, where clear tendencies of optimization results in respect to tuning parameters α and β can be seen. While an approach considering only epistemic uncertainties delivers overall best results in respect to squared errors, aleatoric uncertainties are out of scope, thus, optimization results lead to less robust outcomes. An approach considering aleatoric and epistemic uncertainties combined by summing them up shows overall worst results and can not compete with the remaining. Our approach, where aleatoric and epistemic uncertainties are considered to deliver different information, depicts that overall good results are achieved with respect to squared errors while keeping aleatoric uncertainty low, thus robust solutions for solving multi-objective inverse problems are provided.

Moreover, our model is directly applicable in an industrial framework where the forging press characteristics are represented by measured sensor data of the aggregate (e.g., velocity over time, forging force over time, forging force over the forming path, etc.), which can be used in an appropriately designed FEM simulation for uncertainty propagation analysis and, moreover, for surrogate model training.

7. Conclusions

In this work, we present a GP based reduced-order surrogate model approach with a novel multi-objective target vector optimization strategy to obtain more robust optimization results by concerning aleatoric and epistemic uncertainties. Evaluation on a classic hot metal forming use case, preforming an Inconel 625 forging billet on a self-designed forging press,

depicts the advantages of our approach compared to baselines. Our major findings include that our surrogate model produces fast results (over 3000 times faster) compared to FEM simulation, with a calculated loss of accuracy and information. Moreover, active learning can be used directly with our model to make optimal use of computational resources, and solving inverse problems leads to robust optimization results, i.e., finding results close to a defined objective while keeping aleatoric uncertainty low. With our work, we pave one promising way for faster and more realistic simulation and optimization methods.

In future work, we will evaluate our GP based surrogate model and multi-objective optimization strategy on manufacturing process use cases concerning other domains, with real sensor data describing the characteristics of a manufacturing process. Additionally, we will research other Bayesian machine learning and deep learning models as components instead of GP in our surrogate model approach. Moreover, we will experiment with further active learning approaches.

Author Contributions: Conceptualization, J.G.H. and B.C.G.; methodology, J.G.H. and B.C.G.; software, J.G.H.; validation, J.G.H., B.C.G. and R.K.; formal analysis, J.G.H.; investigation, J.G.H., B.C.G. and R.K.; resources, J.G.H. ; data curation, J.G.H.; writing—original draft preparation, J.G.H.; writing—review and editing, J.G.H., B.C.G. and R.K.; visualization, J.G.H.; supervision, B.C.G. and R.K.; project administration, J.G.H., B.C.G. and R.K.; funding acquisition, J.G.H. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yang, X.S.; Koziel, S.; Leifsson, L. Computational optimization, modelling and simulation: Recent trends and challenges. *Procedia Comput. Sci.* **2013**, *18*, 855–860. [CrossRef]
2. Krige, D.G. A statistical approach to some basic mine valuation problems on the Witwatersrand. *J. S. Afr. Inst. Min. Metall.* **1951**, *52*, 119–139.
3. Makarova, A.; Usmanova, I.; Bogunovic, I.; Krause, A. Risk-averse Heteroscedastic Bayesian Optimization. In Proceedings of the Thirty-Fifth Conference on Neural Information Processing Systems, Virtual, 6 December 2021.
4. Binois, M.; Gramacy, R.B. hetGP: Heteroskedastic Gaussian process modeling and sequential design in R. *J. Stat. Softw.* **2021**, *98*, 1–44. [CrossRef]
5. Tran, T.; Stough, J.V.; Zhang, X.; Haggerty, C.M. Bayesian Optimization of 2D Echocardiography Segmentation. In Proceedings of the 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 13–16 April 2021; pp. 1007–1011.
6. Roberts, S.; Kusiak, J.; Liu, Y.; Forcellese, A.; Withers, P. Prediction of damage evolution in forged aluminium metal matrix composites using a neural network approach. *J. Mater. Process. Technol.* **1998**, *80*, 507–512. [CrossRef]
7. Loghin, A.; Ismonov, S. Augmenting generic fatigue crack growth models using 3D finite element simulations and Gaussian process modeling. In *Pressure Vessels and Piping Conference*; American Society of Mechanical Engineers: New York, NY, USA, 2019; Volume 58936, p. V002T02A004.
8. Ming, W.; Zhang, G.; Li, H.; Guo, J.; Zhang, Z.; Huang, Y.; Chen, Z. A hybrid process model for EDM based on finite-element method and Gaussian process regression. *Int. J. Adv. Manuf. Technol.* **2014**, *74*, 1197–1211. [CrossRef]

9. Su, G.; Peng, L.; Hu, L. A Gaussian process-based dynamic surrogate model for complex engineering structural reliability analysis. *Struct. Saf.* **2017**, *68*, 97–109. [CrossRef]

10. Guo, M.; Hesthaven, J.S. Reduced order modeling for nonlinear structural analysis using Gaussian process regression. *Comput. Methods Appl. Mech. Eng.* **2018**, *341*, 807–826. [CrossRef]

11. Hu, X.; Li, Y.; Zhao, Z.; Liu, C.; Salonitis, K. Residual stresses field estimation based on deformation force data using Gaussian Process Latent Variable Model. *Procedia Manuf.* **2021**, *54*, 279–283. [CrossRef]

12. Yue, X.; Wen, Y.; Hunt, J.H.; Shi, J. Active learning for gaussian process considering uncertainties with application to shape control of composite fuselage. *IEEE Trans. Autom. Sci. Eng.* **2020**, *18*, 36–46. [CrossRef]

13. Ortali, G.; Demo, N.; Rozza, G. Gaussian process approach within a data-driven POD framework for fluid dynamics engineering problems. *arXiv* **2020**, arXiv:2012.01989.

14. Venkatraman, A.; de Oca Zapiain, D.M.; Lim, H.; Kalidindi, S.R. Texture-sensitive prediction of micro-spring performance using Gaussian process models calibrated to finite element simulations. *Mater. Des.* **2021**, *197*, 109198. [CrossRef]

15. Lee, T.; Bilionis, I.; Tepole, A.B. Propagation of uncertainty in the mechanical and biological response of growing tissues using multi-fidelity Gaussian process regression. *Comput. Methods Appl. Mech. Eng.* **2020**, *359*, 112724. [CrossRef] [PubMed]

16. Brevault, L.; Balesdent, M.; Hebbal, A. Overview of Gaussian process based multi-fidelity techniques with variable relationship between fidelities, application to aerospace systems. *Aerosp. Sci. Technol.* **2020**, *107*, 106339. [CrossRef]

17. Civera, M.; Boscato, G.; Fragonara, L.Z. Treed gaussian process for manufacturing imperfection identification of pultruded GFRP thin-walled profile. *Compos. Struct.* **2020**, *254*, 112882. [CrossRef]

18. Abdelfatah, K.; Bao, J.; Terejanu, G. Environmental Modeling Framework using Stacked Gaussian Processes. *arXiv* **2016**, arXiv:1612.02897.

19. Mao, J.; Wang, H.; Li, J. Bayesian finite element model updating of a long-span suspension bridge utilizing hybrid Monte Carlo simulation and kriging predictor. *KSCE J. Civ. Eng.* **2020**, *24*, 569–579. [CrossRef]

20. Tapia, G.; Khairallah, S.; Matthews, M.; King, W.E.; Elwany, A. Gaussian process-based surrogate modeling framework for process planning in laser powder-bed fusion additive manufacturing of 316L stainless steel. *Int. J. Adv. Manuf. Technol.* **2018**, *94*, 3591–3603. [CrossRef]

21. Yu, Z.; Shi, X.; Zhou, J.; Huang, R.; Gou, Y. Advanced prediction of roadway broken rock zone based on a novel hybrid soft computing model using Gaussian process and particle swarm optimization. *Appl. Sci.* **2020**, *10*, 6031. [CrossRef]

22. Lee, S.H. Optimization of cold metal transfer-based wire arc additive manufacturing processes using gaussian process regression. *Metals* **2020**, *10*, 461. [CrossRef]

23. Saul, A.D.; Hensman, J.; Vehtari, A.; Lawrence, N.D. Chained gaussian processes. In Proceedings of the Artificial Intelligence and Statistics, PMLR, Cadiz, Spain, 9–11 May 2016; pp. 1431–1440.

24. Binois, M.; Gramacy, R.B.; Ludkovski, M. Practical heteroscedastic gaussian process modeling for large simulation experiments. *J. Comput. Graph. Stat.* **2018**, *27*, 808–821. [CrossRef]

25. Dai Nguyen, T.; Gupta, S.; Rana, S.; Venkatesh, S. Stable bayesian optimization. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*; Springer: Cham, Switzerland, 2017; pp. 578–591.

26. Huang, C.; Ren, Y.; McGuinness, E.K.; Losego, M.D.; Lively, R.P.; Joseph, V.R. Bayesian optimization of functional output in inverse problems. *Optim. Eng.* **2021**, *22*, 2553–2574. [CrossRef]

27. Uhrenholt, A.K.; Jensen, B.S. Efficient Bayesian optimization for target vector estimation. In Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics, PMLR, Naha, Japan, 16–18 April 2019; pp. 2661–2670.

28. Snelson, E.; Rasmussen, C.E.; Ghahramani, Z. Warped gaussian processes. *Adv. Neural Inf. Process. Syst.* **2004**, *16*, 337–344.

29. Plock, M.; Burger, S.; Schneider, P.I. Recent advances in Bayesian optimization with applications to parameter reconstruction in optical nano-metrology. *Proc. SPIE* **2021**, *11783*, 117830J.

30. Hüllermeier, E.; Waegeman, W. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Mach. Learn.* **2021**, *110*, 457–506. [CrossRef]

31. Williams, C.K.; Rasmussen, C.E. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006; Volume 2.

32. Burbidge, R.; Rowland, J.J.; King, R.D. Active learning for regression based on query by committee. In Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning, Birmingham, UK, 16–19 December 2007; Springer: Berlin/Heidelberg, Germany, 2007; pp. 209–218.

33. Wu, D.; Lin, C.T.; Huang, J. Active learning for regression using greedy sampling. *Inf. Sci.* **2019**, *474*, 90–105. [CrossRef]

34. Meka, R.; Alaeddini, A.; Oyama, S.; Langer, K. An active learning methodology for efficient estimation of expensive noisy black-box functions using Gaussian process regression. *IEEE Access* **2020**, *8*, 111460–111474. [CrossRef]