Noname manuscript No.
(will be inserted by the editor)

Meta-model based optimization of hot rolling processes in the metal industry

Christian Jung ·
Martin Zaefferer ·
Thomas Bartz-Beielstein ·
Günter Rudolph

the date of receipt and acceptance should be inserted later

Abstract To maximize the throughput of a hot rolling mill, the number of passes has to be reduced. This can be achieved by maximizing the thickness reduction in each pass. For this purpose, exact predictions of roll force and torque are required. Hence, the predictive models that describe the physical behavior of the product have to be accurate and cover a wide range of different materials.

Due to market requirements a lot of new materials are tested and rolled. If these materials are chosen to be rolled more often, a suitable flow curve has to be established. It is not reasonable to determine those flow curves in laboratory, because of costs and time. A strong demand for quick parameter determination and the optimization of flow curve parameter with minimum costs is the logical consequence. Therefore parameter estimation and the optimization with real data, which were collected during previous runs, is a promising idea. Producers benefit from this data-driven approach and receive a huge gain in flexibility when rolling new materials, optimizing current production, and increasing quality. This concept would also allow to optimize flow curve parameters, which have already been treated by standard methods. In this article, a new data-driven approach for predicting the physical behavior of the product and setting important parameters is presented. We demonstrate how the prediction quality of the roll force and roll torque can be optimized sustainably. This offers the opportunity to continuously increase the workload in each pass to the theoretical maximum while product quality and process stability can also be improved.

Keywords flowcurve · Kriging · meta-model · metal · hot rolling

FACULTY FOR COMPUTER AND ENGINEERING SCIENCES
Cologne University of Applied Sciences,
51643 Gummersbach, Germany
firstname.lastname@fh-koeln.de

FACULTY FOR COMPUTATIONAL INTELLIGENCE
TU Dortmund University,
44227 Dortmund, Germany
Guenter.Rudolph@tu-dortmund.de

1 Introduction

The complex process of hot rolling requires very accurate physical models. Several different physical models are used in hot rolling mills. These models include the slab or ingot heating at the furnace, where the thermal behavior is modeled, as well as the rolling process itself. The task of the different models is to predict the thermal and physical behavior of the product and set important parameters for achieving maximum product throughput while increasing the quality of the final product. There are several process and product parameters which play important roles. They are independent of the plant type and applicable to steel and aluminum hot mills. Some of them, for example plant geometries or drive parameters, remain constant during rolling. Other parameters may vary during rolling but are not dominated by the material, e.g., the maximum possible thickness reduction depends mainly on the actual thickness and the work rolls currently installed. The maximum redcution is of course also dependent from the friction between the rolls and the material but this effect is not as huge as the geometrical limitations. Parameters, which depend on the material of the product, are usually hard to optimize, because (i) the number of different materials is steadily increasing due to market requirements and (ii) measurements of the process are only indirectly correlated to the material.

The most important parameters for the prediction of the roll force are the flow curve parameters of each material. A more efficient method for optimizing these parameters is necessary to increase the flexibility and reduce the cost of the rolling process.

A recent approach uses artificial intelligent techniques for optimization of shape rolling sequences [17]. Especially in the field of cold rolling, several methods for simulation and optimization were published [18, 24]. There were also studies, which were based on *finite element methods* [25]. In this paper we propose meta-model based optimization strategies for the task of hot rolling mill flow curve parameter optimization. Meta models, also referred to as surrogate models, simplify

the simulation optimization as the run times are generally much shorter than the original function evaluations [1, 16] and are a proven strategy for global optimization [15]. The results are compared to classical optimization strategies. Hence, this paper addresses the following research questions:

(R-1) Can the rolling process benefit from meta-model based optimization?

(R-2) How time- and cost efficient is this approach in comparison to the established industry procedures?
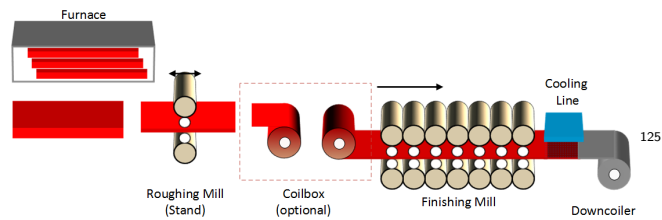
This paper is structured as follows. Section 2 introduces technical terms and fundamental principles of hot rolling and the related parameters. It describes the current state-of-the-art approach in industry. Problems related to the rolling process are described in Sec. 3. Our methodology is detailed and compared to existing approaches in Sec. 4. The experimental setup is described in Sec. 5. Results are presented in Sec. 6. The paper concludes with a discussion in Sec. 7.

2 Hot Rolling

2.1 Fundamentals and Technical Terms

Before describing some of the main aspects, the fundamental technical terminology will be introduced. Figure 1 shows a common hot strip mill for steel. The process in general is very similar to an aluminum mill. The work flow of the process is from the left side to the right side. The main components are a reheating furnace, a reversing roughing mill, a continuous finishing mill, a cooling line, and a downcoiler. The coilbox between roughing mill and finishing mill is more or less optional. It is used to achieve better temperature profiles and allows a more compact rolling mill.

The process starts with the charging of the furnace. Here, slabs which are usually at room temperature are charged and reheated to temperatures around 1200 deg C for steel mills and 500 deg C for aluminum mills.

Fig. 1 Hot Rolling mill for steel with coilbox. The roughing mill consists of one horizontal stand with four rolls, the so-called quarto stand and an optional vertical rolling stand with two rolls, the so-called edger (not shown in the figure).

The slab geometries may vary. Usually, they have an *input thickness*, h_{init}, between 200 mm and 300 mm, width of 600 mm to 2500 mm and length between 3 m and 10 m for steel production. For aluminum, the thickness after discharge is commonly around 600 mm because the temperature loss of aluminum during rolling is much less than for steel. After discharging, the first major process is to reduce the thickness of the slab by 30 mm to 40 mm. This is done in a so-called *roughing mill* (RM). The roughing mill in a conventional hot strip mill consists of one horizontal stand with four rolls, which is then called quarto stand and an optional vertical rolling stand with two rolls, the so-called edger (not shown in the figure). The edger is optional and has the task to reduce the width of the slab and to improve the shape of the slab especially at both ends.

The reduction from the initial thickness h_{init} down to the target thickness h_{target} is done in several deformation steps which are called *passes*. In each of these deformation steps, the thickness of the slab will be continuously reduced until the target thickness is attained. The reduction has to be split to several passes because the feasible reduction in one pass is limited. The deformation in the roughing mill can be done in both operating directions. Hence, each pass changes the direction of movement of the slab. The total number of passes has to be odd, since the slab has to be moved to the next process step. After rolling in the roughing mill, the slab is transferred in the direction of the finishing mill. If a coilbox is used, the material is coiled first and then directly uncoiled to start the rolling in the finishing mill. Here, the product is rolled in several stands to the final thickness (specified by by the customer) and is directly cooled afterwards. The additional cooling line is only used for steel mills. Finally, the product is coiled in the downcoiler.

One important quality criterion for the final product is the deviation of the actual thickness from the target thickness h_{target}. The thickness deviation at the head of the product is typically a result of the roll force prediction accuracy of the physical model. The head of the product is the first part which encounters the deformation. In finishing mills we will most often find thickness gauges after the last stand. These are used to control the thickness once the head passes the gauge so the roll force deviation would mainly be responsible for the head thickness. Additionally, the roll force is used as reference for the bending and for profile control and is therefore also very important for the product profile.

Inline control of the thickness is usually not installed in other mill types such as plate mills. This is because the total length of the product is much smaller than in mills with coil production, where the total length

might accumulate to more than 1000m. In that case, a feedback of the measured thickness during rolling can be used to adjust the roll gap and therefore the final thickness. For mill types without inline control, it is crucial to improve the roll force prediction in order to minimize the thickness deviation.

Another aspect is the throughput of the mill, which should always be maximized. To maximize the throughput, the number of deformation steps, i.e. the number of passes, should be minimized without violating other constraints. Two of the most limiting parameters when trying to increase the reduction in each pass are (i) the maximum roll force capability and (ii) the maximum roll torque capability of the stand. Therefore, it is essential to have good models for the prediction of roll force and torque.

As mentioned above, deviations in roll force prediction will also affect the thickness and therefore the quality of the product. Because current market requirements cover a very wide range of materials and geometries it is important to increase the model quality for the roll force prediction for all products which may be rolled on these mills. Ideally, the roll force prediction is completely independent of the geometry and other parameters and will only depend on the quality of the material parameters. Each material is usually classified according to its chemistry. A material database stores mechanical and thermophysical parameters for the description of the different properties of each class. These parameters are used for the prediction of behavior during the deformation process and are therefore of major importance to the rolling process.

2.2 Flow Curve and Roll Force Model

The flow curve parameters are most relevant for the roll force model. The flow curve expresses the material resistance during plastic deformation in dependence on the chemistry, the temperature, the deformation and the deformation rate. The *deformation* φ, also called effective, logarithmic deformation or true strain is expressed by:

$$\varphi = \ln \frac{h_0}{h_1}.$$

Here, h_0 is the input thickness and h_1 is the output thickness of the pass. The first flow curve formulas were developed by Geleji and Ekelund around 1950. These formulas were only linearly dependent on the temperature and only valid for standard low carbon steel [10]. Afterwards several other formulas were developed with polynomial components and exponential terms which

also took into account the deformation and deformation rate. While the first formulas were only valid for some low alloyed carbon steels, Hajduk developed formulas which were also valid for some medium and high alloyed carbon steels [9],[8]. A good overview and description of the different flow curves can be found in [10, 12, 26]. These formulas model the *deformation resistance*, k_f, in dependence on the deformation φ, the *deformation rate*, $\dot{\varphi}$, and the *temperature*, ϑ. The *deformation resistance* or flow stress expresses the stress which is needed to sustain a plastical deformation. In general, the deformation rate can be formulated as:

$$k_f = A K_\varphi K_\vartheta K_{\dot{\varphi}},$$

where $A \in \mathbb{R}_+$ is a constant factor and the terms $K_{(\cdot)}$ represent functions of the corresponding variables φ, $\dot{\varphi}$, and ϑ, respectively. The most common model for these formulas was developed by Hensel and Spittel [10]. It was extended at the University of Freiberg. Thus, these extensions are called *Freiberger Approach*. The extended versions of this flow curve model gives a better approximation of the flow stress within high deformation grades. Some of the available flow curve models, which are typically used in process models for hot rolling, are presented in Table 1. Their corresponding equations read as follows.

$$k_f = k_{f,0} A_0 A_1 e^{m_1 \vartheta} A_2 \varphi^{m_2} e^{\frac{m_4}{\varphi}} A_3 \dot{\varphi}^{m_3} \tag{1}$$

$$k_f = A_0 e^{m_1 \vartheta} \varphi^{m_2} e^{\frac{m_4}{\varphi}} \dot{\varphi}^{m_3} \tag{2}$$

$$k_f = A_0 e^{m_1 \vartheta} \varphi^{m_2} e^{\frac{m_4}{\varphi}} (1 + \varphi)^{m_5 \vartheta} e^{m_7 \varphi} \dot{\varphi}^{m_8 \vartheta} \tag{3}$$

$$k_f = k_{f,0} A_1 e^{m_1 \vartheta} A_2 \varphi^{m_2} A_3 \dot{\varphi}^{m_3} \tag{4}$$

The multipliers A_i ($i = 0, 1, 2, 3$) can be reduced to one parameter, A. The parameters m_j ($j = 1, 2, \ldots, 8$) are defining the exponential behavior of the materials in dependence of the temperature ϑ, the deformation φ, and the deformation rate $\dot{\varphi}$. The parameters ϑ, φ, and $\dot{\varphi}$ are defining the working point in each deformation. The value $k_{f,0}$ used in the equations (1) and (4) is the *basic deformation* and is calculated by empirical formulations based on the chemistry. Each material is classified according to its chemistry and gets its own parameter set $\mathcal{M}$ with parameter values m_1 to m_8 and A_0 to A_3, respectively. Usually, there is one parameter-set $\mathcal{M}$ for each material, which is then valid for a specific equation only. Besides the parameter values for the models also the valid region of these parameters is stored. Summarizing, the parameters $k_f, \varphi, \dot{\varphi}$, and ϑ, the multiplicators A_i, m_j, ($i = 0, 1, 2, 3; j = 1, \ldots, 8$), and related functions $k_{f,0}, K_\varphi, K_{\dot{\varphi}}, K_\vartheta$, are used.

Nowadays, hundreds of different materials are known. The parameter k_f is almost linearly correlated with the

Table 1 Overview of typical equations for the regression of the material flow stress. The multipliers A_i are reduced to one parameter, A. Parameters m_j are defining the exponential behavior of the materials in dependence of the temperature ϑ, the deformation φ or the deformation rate $\dot{\varphi}$ which defines the working point. Parameter $k_{f,0}$ is defined by a simple equation based on the chemistry. Entries in the column "Equation" refers to the equations defined on p. 3.

Eq.	Name	#Params	Parameter List
(1)	Freiberg 1	5	A, m_1, m_2, m_3, m_4
(2)	Freiberg 4	5	A, m_1, m_2, m_3, m_4
(3)	Freiberg 8	7	$A, m_1, m_2, m_4, m_5, m_7, m_8$
(4)	Hensel Spittel	4	A, m_1, m_2, m_3

roll force and roll torque. Thus, the model prediction quality and herewith the process stability and product quality are correlated to the parameter-set $\mathcal{M}$. It is therefore important to optimize those parameters in order to increase the model quality and to ensure a stable process with maximum throughput and product quality. A standard procedure for obtaining these parameters is the measurement of the deformation resistance in a laboratory. Those measurements can be used for a regression onto one of the formulas shown in Equations (1) to (4). Of course, other formulas exist, and might be used for regression. Especially when trying to model the deformation of micro alloyed or high alloyed steel or when complex materials with phase transformations should be described these other models might be more suitable.

2.3 Description of the analytical model

The analytical model used for the calculation of the roll force is based on the elementary rolling theory [12, 27]. Some of the limitations of that theory are compensated with correction functions. For example, one of the requirements for the elementary theory is that during each pass there is a complete plastic deformation of the whole material. For products with thickness above 500 mm this is clearly not given. Therefore, a compensation curve, which is empirically determined, is applied. For the calculation of the roll force in each pass the deformation zone is divided into single stripe-like elements and the force balance for each stripe is calculated. The solution yields to the basic differential equation of the plastically deformation theory which was developed by Karman in 1925. When calculating the roll force for one pass, the flow resistance has to be considered. This flow resistance k_f is material dependent and is influenced by the parameter values of φ, $\dot{\varphi}$ and ϑ. Furthermore, the forces induce temperature into the material so the calculation of the next pass depends on the pre-

vious passes. Optimizing the flow curve by analyzing the rolling results is not sufficient. If the parameter of the flow curve changes, the whole process has to be simulated again and then the calculated roll forces based on the new flow curve can be compared with the original feedback, i.e., measurements of roll force, torque, temperature, and speed. Additionally, it is also not sufficient to optimize the result of a single product because the parameters φ, $\dot{\varphi}$ and ϑ may not vary enough to achieve stable results. Therefore it is preferable to consider a campaign with a wide variation of product geometries, temperature ranges, and deformations.

2.4 Standards in Industry

Currently, the available concepts for the optimization of flow curve parameters are mostly dealing with determination of those parameters in laboratory rather than optimizing those parameters with real process data. Traditionally, the parameters are measured with small samples of one piece in laboratory and are then generalized for every material which is close to the sample in terms of material composition. Some companies are modifying the flow curve parameters with linear models. That is, they are determining the prediction accuracy of their model and are varying some of the influence parameters. Most of the research in this area is on the development of suitable flow curve equations especially for high and micro alloyed steel [11, 19, 28] rather than using a data driven approach for the optimization of those parameters.

3 Problem Description

According to our best knowledge, flow curve parameter determination in laboratory as described in Sec. 2 will require several weeks and costs several thousand Euro for the required materials. Sometimes, this is not affordable and therefore not a suitable way to determine those parameters. Hence, we are looking for a cheaper approach to parameter estimation.

Due to their highly nonlinear behavior, the flow curve equations cannot be solved directly. Furthermore, a different roll force would result in a different temperature balance of the product and thus the temperature in the next pass differs from the original calculation. Because these aspects cannot be neglected, we have to simulate a whole scenario when testing new parameter sets for the flow curve of a specified material. The calculations of the roll forces and roll torques within this simulation are afterwards compared with the measurements to get a quality criterion for the new parameter

set. The detailed description of the simulation scenario is presented in Chapter 4.

It is important that the simulation scenario behaves in the same way as the online process. Therefore, the simulation uses feedback of the measured speed, the reduction, and temperature to calculate the new settings. This enables the simulation to achieve the same working point as in the online process. Another problem can be the amount of data. The simulation of a whole batch where only one material group was rolled consists of thousands of different deformation steps and will therefore be highly expensive in terms of simulation time. Several optimization algorithms require bound constraints of the optimized parameters. In our case, parameters A and m_i are dependent on each other. The only limitation which can be set is a plausible region for the resulting value k_f of the basic deformation. In hot mills, the maximum basic deformation value for k_f is usually below $300\,\frac{N}{mm^2}$, but always positive. Then, for a given maximum working range of the deformation, deformation rate, and temperature, the feasibility of the parameter set can be tested.

Due to the fact that every company has usually its own classification system it might be that materials which are grouped together in one company are separated in other companies. In this case, the optimization, which has been done in the first company cannot be directly used for other companies and has to be renewed every time.
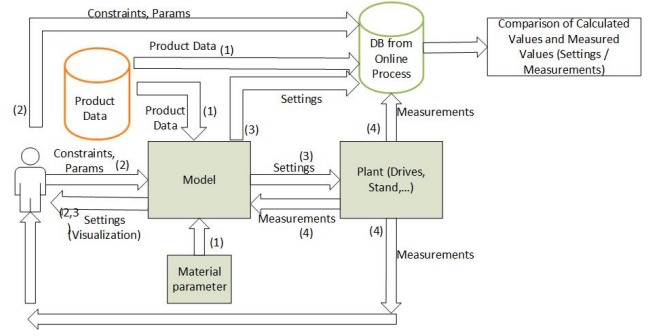
Summarizing, it is desirable to optimize the process in order to
- reliably estimate valid flow curves,
- reduce lab costs,
- save time,
- determine parameters in their working environments, and
- make the process more flexible and adapt to new (material) changes quickly.
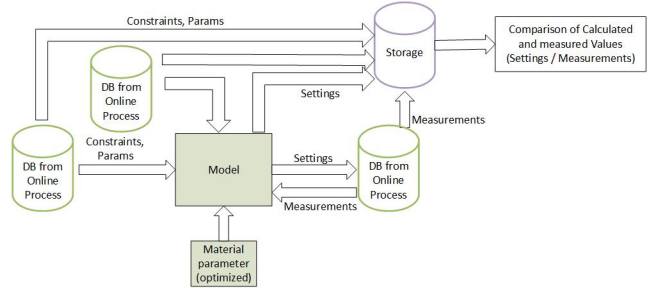
4 Methodology

4.1 Simulation environment

The environment of the online process is shown in Fig. 2. First of all the model gets information about the product which includes initial geometry data, discharging temperature data, and information about the chemistry of the product. The discharge temperature is an initial temperature field for the product. One part of the rolling model calculates the temperatures losses during the whole process. Finally, the important parameter ϑ is a result of the temperature losses from discharge to that point of the process. Furthermore, target data is

Fig. 2 Model environment in the real-world process for each pass: The product and customer data such as material description, initial and target geometry (1) are combined with operator data (2) and are fed to the model which calculates all required settings (3) for the next rolling pass (setup for next pass). After rolling of this pass the model gets feedback (4) of the just rolled pass and combines this information for the recalculation of the previous pass and for the next calculations (2-4). Additionally, changes from the operator for the next pass are send to the model. The product and customer data are only product and not pass dependent and may only be send once. To enable a simulation of this process every in- and output of the model is stored in a database.

Fig. 3 Model environment for the offline simulation. The data which has been collected in the real word process is send to the model which calculates a new setup for the next pass. This setup may be different from the original one. But because of the fact that we also store the feedback from the drives and gauges the model will receive also the original setting and recalculates the pass as it really has been rolled.

also coming from the customer. Both data can be seen as constant and are denoted with N_{const}, i.e., the operator cannot change them as they are part of the production planning system which handles the orders of the mill owner. Afterwards, the N_{const} is used to calculate the first settings for the mill and this result is shown to the operator. The setting consists of roll gap settings, speed settings, geometry, temperature and time calculations. With this data the first planned settings N_{Set}, which also include the parameter φ, ϑ and $\dot{\varphi}$, are calculated and are shown to the operator. The operator can interfere and modify the way this product is rolled. This is referred to as *rolling strategy*. This *rolling strategy* defines how a product is rolled which includes rolling speed, number of passes, deformations, speed settings,

possible rolling breaks and much more. These strategies may also be specific constraints like absolute reduction, deformation φ, force, torque, but also other restrictions to the process like drive limitations.

If the operator is satisfied with the settings N_{Set} calculated by the model, the settings are sent to the plant where the first pass is rolled. After this pass, all measurements collected during rolling which include forces, torques, speeds, temperatures, gap settings, delay times, and several more data, are sent to the model and the database but are also shown to the operator. Now, the operator and also the model can adapt the settings for the next passes and react to any unexpected behavior of the mill. Usually, no big changes are made by the operator and the rolling of the further passes is started directly. Again, for each pass the settings are sent to the plant and the feedback of measurements is received from the plant. For analysis and offline simulations all inputs and outputs are stored in a database.

This database is the basis of the offline simulation shown in Figure 3. Here, we feed the same model which was used in the online process with the data stored during the real-world process. Therefore we can guarantee that the model reacts in the same way as it would react in the rolling process. The data, which was coming from the operator and the geometry data are taken from the database. Hence, the model will not recognize if it is used for an offline calculation or for an online scenario. Although it may calculate different settings for the process, it will receive the original feedback from the plant and calculate everything based on the original settings.

For a product with 19 passes the model is triggered 20 times. The first trigger creates the initial setup and all other triggers are feedbacks for the 19 passes with which the original products were rolled. The only difference are the parameters used for the calculation of the roll force. Therefore, we have a calculated and measured value for the force and torque for each product and pass for every run. After each simulation run, the parameters may be changed and results of different parameters may be compared. This enables the optimization of the flow curve parameters. Minimization of the *Root Mean Square Error* (RMSE) of the predicted roll force is the optimization objective.

4.2 Surrogate Modeling

If the simulation runs or the original problem in general are very expensive in terms of evaluation it would be very time consuming to perform parameter optimization on those original scenarios. Therefore, we use a surrogate-model based optimization approach. Surrogate models are supposed to replace the original, expensive simulation model and are expected to be cheaper to calculate.

The analytical models, as introduced in Sec. 2.2, are cheaper to evaluate in comparison to the finite element method models. In this paper, the term surrogate model is used to describe data driven models, which are built from an analytical model. Therefore, the simulation runs on the analytical model with a specified set of products is the expensive model. The products which are simulated usually belong to the same material group and have previously been rolled in a series on a real rolling mill. Each product is calculated as it would be done during rolling. Therefore, time delays which occurred during the real production are also taken into account.

Our data is sourced from a reversing aluminum hot mill, which has typically around 19 passes. Hence, we have more than 20 calculations for each product because the model is triggered after finishing each pass. The data of each pass is sent to the model, which may react on unpredicted circumstances.

In general, data-driven surrogate models can be any kind of models, e.g., artificial neural networks, linear models, Kriging, random forest and others. A detailed overview of surrogate model based numerical optimization is presented by Jin [13] and Jones [15].

One framework for surrogate-model based optimization is *sequential parameter optimization* (SPO) [3]. SPO combines methods from classical DoE and modern *Design and Analysis of Computer Experiments* (DACE) [2, 4] based on Kriging models.

Algorithm 1 presents the pseudo code of SPO, as adapted for the application of hot mill parameter optimization. Note, that we will use the notation $\left(x^{(i)}, y^{(i)}\right)$ for the data from the i-th pass which is passed to the surrogate model. During the first stage of experimentation, SPO explores the search space of the optimization problem A, which is treated as a black box. A set of input design points $\mathbf{x}$ is passed to A. Usually these are created by a space filling design, e.g. Latin hypercube sampling. Each call of the objective function produces some output y regarding its performance.

SPO now tries to determine a functional relationship between $\mathbf{x}$ and y. SPO thus uses a model $Y(\mathbf{x})$ as surrogate for the hot mill simulation model A. As mentioned above, the chosen model type is Kriging Kriging is frequently used for surrogate-model based optimization, because it provides a powerful and flexible predictor. It also provides an estimate of the variance or error of each prediction. The observations are interpreted as realizations of a stochastic process. A gaussian kernel is used to model the correlation between observations [23].

Algorithm 1: SPO-based hot mill simulation tuning.

```
// phase 1, collect initial knowledge about the
   optimized process:
```
1 let A be the hot mill simulation model we want to tune;

2 generate an initial design DES = $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\}$ of n parameter vectors;

3 let $k = k_0$ be the initial number of replications for determining estimated responses;

4 foreach $\mathbf{x} \in DES$ **do**

5   run A with $\mathbf{x}$ to determine the estimated response y of $\mathbf{x}$;

```
// phase 2, building, using and improving a
   surrogate model:
```
6 while *stop criteria not reached* **do**

7   build surrogate model $Y(\mathbf{x})$ based on DES and $\{y^{(1)}, \ldots, y^{(|\text{DES}|)}\}$;

8   optimize the model w.r.t some cost function and constraints, thus produce a set DES' of d new parameter vectors ;

9   run A with each $\mathbf{x} \in$ DES' to determine the response;

10   extend the design by DES = DES $\cup$ DES';

```
// phase 3, final exploitation and fine tuning:
```
11 use local optimizer for the best p parameter sets $\mathbf{x}_{1\ldots p} \in$ DES, without constraints

---

In the sequential improvement loop SPO optimizes the surrogate model $Y(\mathbf{x})$ over the considered space of input variables by means of a cost function. Once the new set of design points DES' has been selected, the required evaluations of DES' are performed. Based on DES', the surrogate model $Y(\mathbf{x})$ is updated.

In step 8 of Algorithm 1, a search on the surrogate model is performed. Here, the constraints of the mill parameterization problem have to be considered. As the constraints are not expensive to evaluate, they are evaluated together with the surrogate model itself. For the inequality constrained optimization, we use the popular method developed by Powell [21, 22], which does not require any derivatives of the objective function to be available. During this optimization step 8, the next point $\mathbf{x}$ to evaluate in the sequential loop of SPO is determined. For expensive, global, black-box optimization Jones [14] introduced *efficient global optimization* (EGO). EGO exploits the information given from a Kriging model, i.e., the predicted mean and variance, to compute the *expected improvement* (EI) of a given solution. EI can hence be used as a cost function during step 8, as an alternative to the predicted value of the Kriging model.

In step 11, the well known downhill simplex algorithm introduced by Nelder and Mead [20] is used to improve the best found results by a local optimization procedure. We choose the downhill simplex implementation in the `nloptr` R package. During local refinement, constraints are disregarded because they no longer play a role in the region of good solutions.

5 Experimental Setup

In our case, the parameter optimization was based on Equation (1). The feasible range was $0 \leq k_f \leq 300$. That is, solutions that result into negative k_f values or k_f values larger than 300 are considered to be infeasible. The usual working point for our test data was in the following range:

$$0 \leq \varphi \leq 0.5, \quad 0 \leq \dot{\varphi} \leq 600, \quad 500 \leq \vartheta \leq 600.$$

With that said, the optimization problem to be solved in this study is defined as follows:

- Parameters to be changed are the flow curve parameter vector $\mathbf{m}$ and the consolidated parameter A of the flow curve.
- The deviation of simulated roll force from the measured roll force is minimized.
- Computational constraints: The evaluations of the objective function is expensive. (see Section 4.2)

The parameters which represent the search space and were subject to optimization in this study are summarized in Table 2.

To evaluate the success of the optimization, the resulting parameter set is compared to a well-known, established parameter set used in practice so far. This old parameter set has been determined by experts according to best knowledge from literature on similar materials.

We have chosen the *SPO toolbox* (SPOT) to conduct the experiments [5]. SPOT itself has parameters as well that are set according to the authors experience:

- The chosen surrogate model is Kriging, based on code by Forrester et al. [7].
  Parameter: `seq.predictionModel.func`.
- The initial design consists of 40 candidate solutions, which are created by Latin Hypercube Sampling (LHS).
  Parameter: `init.design.size`.

Table 2 Upper and lower bounds for the parameter set $\mathcal{M}$ introduced in section 2.2 which was used during the optimization. All parameters are of type FLOAT.

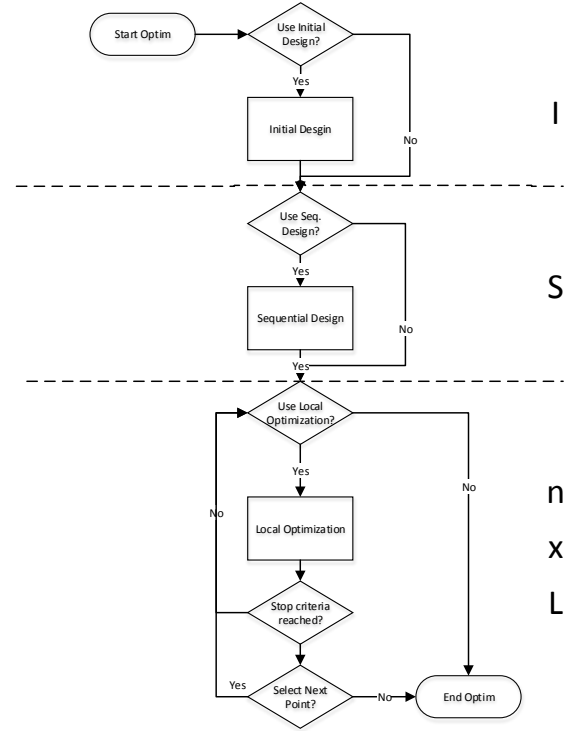| Factor	Low	High
A	0	2
m_1	-0.01	0
m_2	-0.3	0.4
m_3	0	0.2
m_4	-0.1	0.1

Table 3 Spot settings. A detailed description of each parameter can be found in 5

Parameter	Value
seq.predictionModel.func	spotPredictForrester
seq.predictionOpt.func	spotModelOptim
seq.predictionOpt.method	`"NLOPT_LN_COBYLA"`
seq.predictionOpt.budget	1000
seq.predictionOpt.restarts	TRUE
seq.design.size	200
seq.design.new.size	1
init.design.size	40
init.design.repeats	1
auto.loop.nevals	40/50/100
seq.infill	spotInfillExpImp/NA

- Overall, we varied the number of evaluations of the objective function (hot mill simulation) with varying sequential design. We choose the sequential design either consist of 10 evaluations of the objective function or 60 evaluations of the objective function. Therefore we had 50 or 100 evaluations in total.
  Parameter: `auto.loop.nevals`.
- Since the objective function is deterministic each solution is evaluated exactly once.
  Parameter: `init.design.repeats`.
- In each sequential step, one new solution is evaluated on the target function.
  Parameter: `seq.design.new.size`.
- The sequentially created models are optimized by Latin Hypercube Sampling (LHS) and Constrained Optimization by Linear Approximation (COBYLA). The COBYLA implementation from the NLOPT library, included with the `nloptr` R-package is used.
  Parameter: `seq.predictionOpt.method`.
- The sequential step LHS evaluates 200 points with the model.
  Parameter: `seq.design.size`.
- COBYLA has a budget of 1 000 evaluations of the surrogate model, and will restart if it converges prematurely.
  Parameter: `seq.predictionOpt.budget` and `seq.predictionOpt.restarts`.

Table 3 summarizes the these chosen values for each parameter.

6 Results

Figure 4 shows the flow chart of the optimization. Throughout this section the notation of the simulations is $D + n \cdot L$ where D is the sum of the initial design size (I) and the sequential design size (S), n is the number of optimized points in the local optimization and L is the number of evaluations on a single point. Since the ini-

Fig. 4 Flow chart of the simulation presented in this section. First of all an initial design (I) is evaluated followed by an sequential design (S). Afterwards a local optimization (L) is performed.
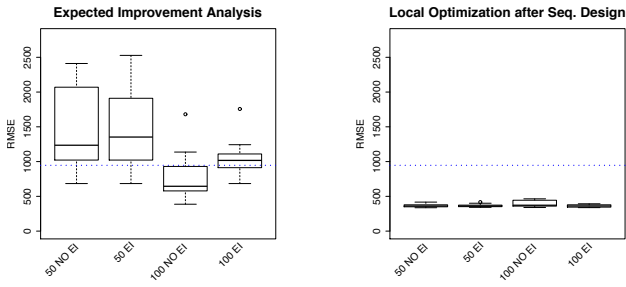
tial design size, if used, is always 40 the size of the sequential design can be directly seen from this notation. When referring to a $50 + 5x100$ simulation the initial design size is 40, the sequential design size is 10 and a local optimization for the best 5 points is performed with 100 evaluations on each point.

6.1 The influence of EI

The first question, which arises when analyzing the simulation results, is whether we can benefit from the expected improvement infill criterion. Therefore, a scenario with initial design size I=40, sequential design size S=10 or S=60 and 10 repeated evaluations was chosen. The experiments with a sequential design size of 60 outperformed the experiments with a sequential design size of only 10. No statistically significant difference between the experiments with EI and without EI can be seen so EI seems not to improve the results even with larger sequential design. In fact it is quite the reverse: The performance of the model-based optimization decreases if EI is used. This is owed to the fact that the total number of evaluations is higher.
The experiments with a sequential design size of 10 had

(a) Analysis of the Kriging simulation with an Initial Design Size of 40 and a sequential design Size of 10 and 60. Shown are the results for the SPOT runs with and without EI based on 10 and 60 evaluations in the sequential design phase. The number in the labels are indicating the total number of evaluations on the original problem.
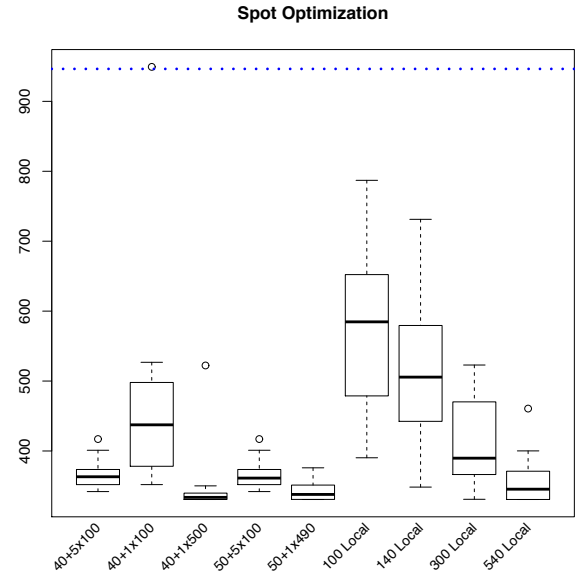
(b) Analysis of the results after local optimization of the previously results found with Kriging. Shown is the best point in each seed after 100 local optimization runs of the 5 best points.

Fig. 5 Comparion of the results for Kriging with and without Expected Improvement after sequential design 5(a) and after the local optimization. The blue dotted line shows the reference value for the original parameter set. 5(b).

a median RMSE of 1235 (no EI) and 1352 (EI) and the experiments with the greater design size of 60 had a median RMSE of 644 (no EI) and 1016 (EI). The results are summarized in Figure 5(a) . To answer the question of the influence of EI it was shown that EI seems to drop or at least not improve the optimization. The major improvement seems to come from the initial design and the local optimization afterwards which is subject of the study in the next section.

6.2 Influence of the local optimization and initial designs

The same comparison as in the previous section was made for the additional refinement of the best-found results with local optimization. Here, the best five parameter sets found by the Kriging-based optimization were selected and optimized with a Downhill Simplex approach. Each of the five parameter sets received a budget of 100 evaluations for further improvement. The local optimization improved the median of the RMSE to below 500 in all experiments whereby only one experiment without the local optimization had a resulting RMSE below 1000. This was the case for an initial design size of 40, a sequential design size of 60 without EI. The local optimization yields the major improvement independent of the budget of the model-based optimiza-

Fig. 6 Comparison of simulation results for an initial design of 40 points with different local optimization afterwards against a pure local optimization with a random starting point. As a reference box also a result for a simulation with a sequential design and local optimization is shown (50+5x100). The boxplots are representing the results for 10 different seeds. The number of parameter evaluations in case of a previous initial design is $40 + 5 \times L$, where L denotes the number of local optimization steps. The blue dotted line shows the reference value for the original parameter set.

Table 4 Numerical values for the boxplots shown in Figure 6. Shown are the number of evaluations in the initial and sequential design and during the local optimization. Box plot parameters listed are minimum, lower quartile (Q1), median, upper quartile (Q3) and maximum value.

Init.	Seq.	Local	Min	Q1	Median	Q3	Max
40		5x100	342	352	363	374	401
40		1x100	352	378	437	498	527
40		1x500	331	331	334	340	350
40	10	5x100	342	352	361	374	401
40	10	1x490	331	331	338	351	376
		100	390	479	585	652	787
		140	348	442	506	580	731
		300	331	366	390	470	523
		540	331	331	345	371	400

tion. The results are shown in the right of Figure 5(a). Thus, the question is whether model-based optimization can be skipped, and whether a local optimization directly after the initial design is more profitable. In addition, it was also checked whether it can be beneficial to choose a completely randomized start points for the local optimization. The results of this investigation are shown in Figure 6. To have a decent comparison the number of evaluations should also be taken into account. The boxplot is showing the statistic for the best

Table 5 Experiments, which were performed to analyze the impact of the initial design and local optimization.

Name	Description	Evaluations
40+5x100:	Initial design of 40 points with 100 local optimization steps of the best 5 points.	540
40+1x100:	Initial design of 40 points with 100 local optimization steps of the best point.	140
40+1x500:	Initial design of 40 points with 500 local optimization steps of the best point.	540
50+5x100:	Initial design of 40 points, model-based optimization with 10 points, and 100 local optimization steps of the best 5 points.	550
50+1x490:	Initial design of 40 points, model-based optimization with 10 points, and 490 local optimization steps of the best point.	540
100 Local:	Random start point with 100 local optimization steps.	100
140 Local:	Random start point with 140 local optimization steps.	140
300 Local:	Random start point with 300 local optimization steps.	300
540 Local:	Random start point with 540 local optimization steps.	540

point produced by each of the 10 random number generator seeds. To analyze the importance of the initial design and the local optimization the experiments from Table 6.2 were performed.
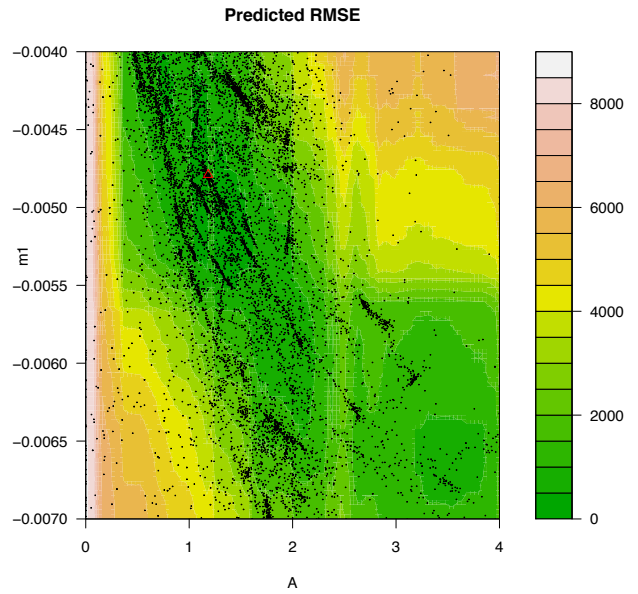
The completely local optimization (with random start point) performs similarly good as the two approaches where the five best points are each optimized locally. The clearly best approach seems to be $(40 + 1 \times 500)$. As a result of the investigation in this section it can be concluded that it is more profitable to optimize just the single best point found in an initial design.

6.3 Impact of the fitness landscape

However, the completely randomized, local approach is not much worse. This indicates that the problem landscape is rather simple. If the landscape would be univariate the optimizations would always result in the same best parameter set independent of the seed. We can see that there are huge differences in the parameter set which are shown in Table 6. This gets more clear if we keep in mind that the first parameter is a multiplicative component and is therefore linearly correlated to the flow resistance. The corresponding contour plot is shown in Fig. 7. While the contour plot depicts several basins in the fitness landscape, they are all of some

Table 6 Parameter comparison for local optimizations with an arbitrary starting point and a budget of 540 evaluations. Only seed 2 stopped earlier, i.e., after 431 evaluations. Shown is the best parameter of each seed with the corresponding RMSE value. The parameters are used to calculate the flow stress according to 1.
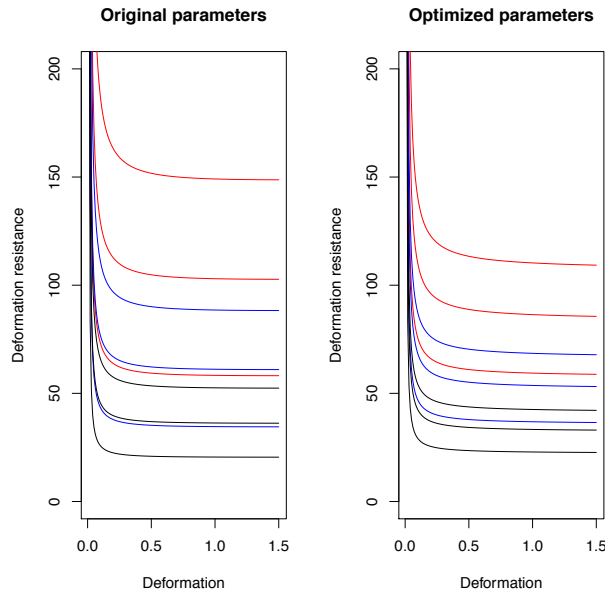
Seed Unit	A	m1 e-03	m2 e-02	m3 e-02	m4 e-02	RMSE kN
1	1.2	-4.8	-0.2	8.2	2.6	331
2	1.2	-4.8	-0.3	8.1	2.6	331
3	0.9	-4.6	-2.7	11.0	2.8	339
4	1.2	-5.0	-5.7	9.8	2.3	335
5	1.0	-5.3	-18.0	14.0	1.8	371
6	0.7	-4.2	-3.2	11.0	2.9	359
7	0.8	-3.5	16.0	5.9	3.9	400
8	1.2	-4.8	-0.5	8.2	2.6	331
9	1.4	-5.5	-6.0	11.0	2.7	352
10	2.9	-5.7	0.8	-0.58	0.95	461

Fig. 7 Contourplot for parameters A and m1. All other variables are kept constant at the same value as for the best case. A random Forest model was created with all available simulation results and is used here for the prediction of the RMSE, denoted by the background coloring. Each black point indicates the result of a simulation. The red triangle marks the simulation with the best RMSE.

considerable size. Note that the contour plots of other parameter combinations result into even simpler landscapes, most giving the impression of being unimodal.

6.4 Significances and plausibility of the findings

We have some knowledge about the shape of the flow curves as well as the correlation to temperature and
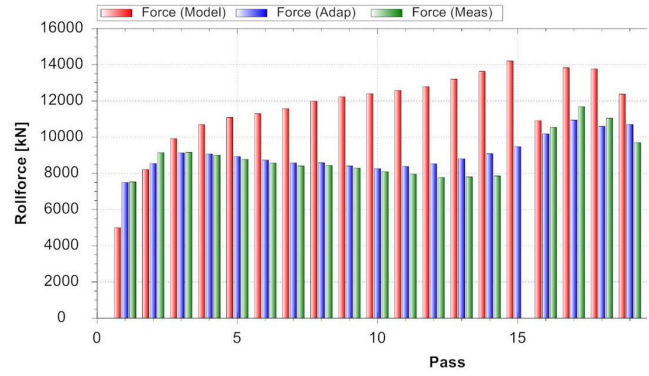
Fig. 8 Deformation plot with the original parameter set and the optimized parameter set. Parameters $\hat{\varphi}$ and ϑ are fixed. For the red lines the temperatures are 300 deg C, the blue lines representing temperature of 400 deg C and the black lines representing temperatures of 500 deg C. $\hat{\varphi}$ has values of 0.1, 10 and 200.

deformation. Thus, we can check if the optimized parameters are describing the expected behavior. The parameter ranges for optimization were chosen in such a way that inverted behavior would have been possible. Consequentially, those parameters have to be plotted and compared to standard parameters. Figure 8 shows a plot of the deformation resistance against the deformation for different parameters of the deformation rate and temperature. It can easily be seen that the general shape of the optimized parameter set is matching the shape of the original parameter set.

6.5 Comparison of the resulting roll force deviations

Figure 9 shows the deviations of the pure model roll force (red), the adapted roll force (blue) and the measured roll force (green) of the original parameter set, before optimization. Figure 10 shows the same for the optimized parameter set. The model roll force is highly correlated with the calculation of the parameter k_f which was described above. The calculation of the roll force was done in a production scenario. According to previous knowledge of the rolling process the system learned the roll force deviation for this material and made an online regression for the correction of the calculation. Furthermore a pass to pass adaption was used to correct the roll force deviation. That is the reason why the

Fig. 9 Roll force of pure model (red), model with online adaption (blue) and measured roll force (green) with original parameter set. The online adaption is typically a simple regression based optimization which learns the difference between model and measured of the last 2 month of production. Additionally the difference of the last pass between prediction and measurement is taken into account. The adaption is used as an additive offset to the model prediction. Because it can be reseted and should only dynamically react on slight deviations from unknown source the correction amount is limited. This limitation can be seen especially in passes 12-14 where we still haven a residual error. The missing measurement in pass 15 is due to a special strategy and problems during the measurement preparation in this pass.
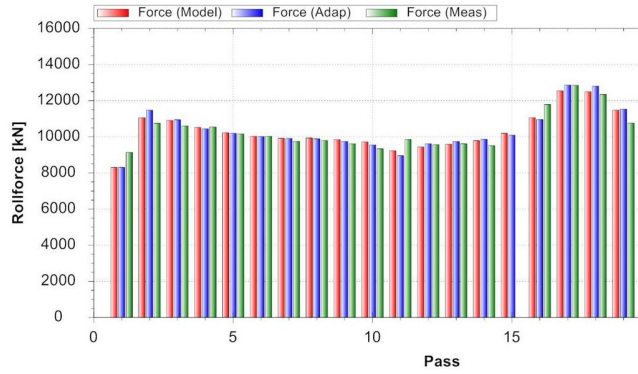
adapted roll force is often close to the measured roll force. Due to limitations to the adaption system some large deviations remain, especially when the model roll force estimate is way off.

Typically, model corrections of up to 30% are tolerated in the online scenario. This limitation is used because something else has to be wrong if calculation errors are exceeding this value. Furthermore, the adaption is reset from time to time. In case of the original parameter set, the deviation were more than 30% and therefore we still have a residual deviation for the rolling force.
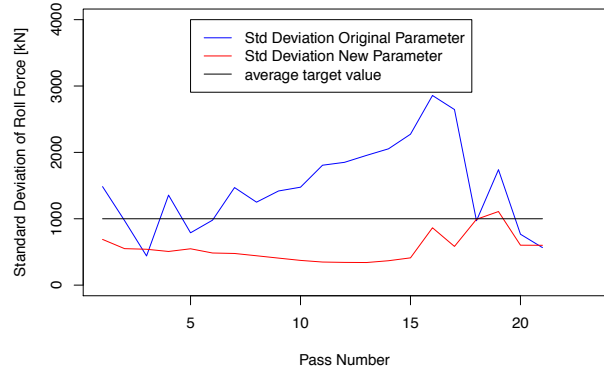
In both figures, Fig. 9 and Fig. 10, the values for the pass number 15 are missing. This is due to a special strategy during rolling of this material. In this early stage of commissioning there were still some difficulties during measurement preparation.

Usually, pure model deviations without any optimization of about 10% are tolerated and deviations of less than 5% are considered to be good.
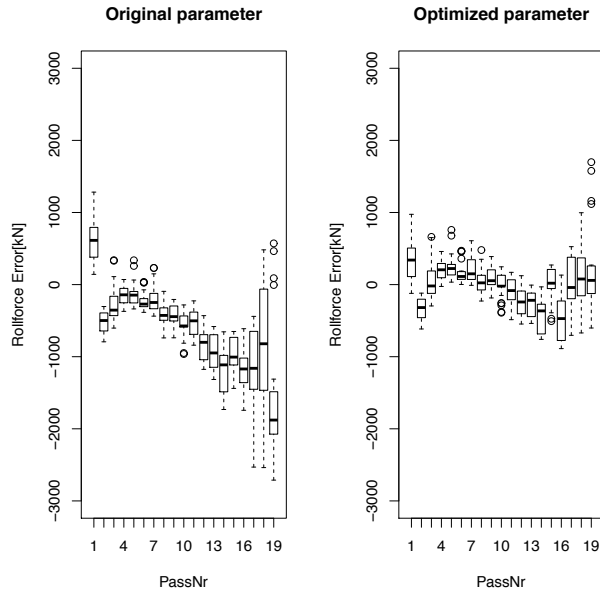
In Figure 11, the roll force deviation of the model with the original, standard parameter set is shown. Next to it, the roll force deviation of the model with the new parameter set is plotted. It can be seen, that the deviation is very good in the first passes but still shows some deterioration in the last passes. This phenomenon accounts for wrong temperature calculation and the process was often delayed which was not correctly modeled during this phase of the process. If these parameters are

Fig. 10 Roll force of pure model (red), model with online adaption (blue) and measured roll force (green) after optimization. Here we have used the parameters of our offline simulation, i.e., the result of the local optimization. This figure clearly demonstrates that the performance of the model has vastly improved and that the online regression could have been switched off.

Fig. 12 Standard deviation of the roll force prediction error for each pass. Shown is the deviation with the original parameter-set and the optimized parameter-set. The target during our simulation is additionally plotted.

optimized in a later stage of the project, the deviation in the last passes will also be reduced to a suitable amount. When the calculated roll force deviates from the actual measurement the automation system is correcting the roll gap position based on the current roll force. This is done in order to compensate for the mill spring. The mill spring is the spring which occurs in the stand when we suddenly have to apply a roll force. This means that the gap between both rolls will be higher and we have to adjust the gap setting to reach a specific thickness. If we consider this opening as a linear function of the roll force and assume a value of $1mm$ at a roll force of $6000kN$ than we would have an error of 0.5 mm when the prediction is 3000 kN away from the measured roll force. When optimizing this prediction, the length of the final product, which is in an acceptable range, can be increased.

During optimization the standard deviation is of major importance and is shown in Figure 12. The aim was a standard deviation which is lower then 1000 kN which we achieved in a very early state during optimization.

Fig. 11 Absolute roll force deviation plotted for each pass. The error is plotted as the difference `Measured Force - Calculated Force`. Since there are too few data for more than 19 passes those data is ignored. The first plot shows the prediction error with the original parameter-set and the second one uses the optimized parameter-set. The drift which is present in the original parameter set is reduced with the optimized parameters. Furthermore, also the standard deviation is reduced with the usage of the optimized parameter.

7 Discussion

This study presented a model-based approach for the optimization of the flow curve parameters used in rolling mills. A combination of both a global and local optimization approaches was used. The data used in the examples was taken from an aluminum mill. The results of the optimization show a significant increase in model prediction quality. This new approach makes it possible to evaluate and optimize the flow curves without con-

ducting time consuming and expensive measurements in a laboratory.

One point of future research of this method is its generalizability. By optimizing those parameters we will also average out all customer specifics like material preparation, furnace heating behavior, alloy description and all plant and measurement based errors. Therefore, a simulation scenario which simulates multiple plants with similar materials should be taken into account when optimizing flow curve parameters.

Another problem might be the pass schedule which was used for the simulation. If we have no or less variance in one of the flow curve representing factors like φ, ϑ or $\dot{\varphi}$ then the optimized coefficient for this factor might be not accurate. Optimization should only be performed for parameters which vary over the whole working point.

If only similar working points in the pass schedule are found, the optimized parameters should be visualized and compared to those used for similar materials. Especially for aluminum, there may be a problem when trying to use a date-driven approach with data which are coming from roughing mills because the temperature variation is rather small. A solution for this problem could be to freeze the coefficient in the flow curve equation for the temperature or use a coefficient for a similar material and optimize only the remaining parameters. This ensures that at least the resulting parameters can be interpreted physically in comparison to other materials. It should be analyzed in further studies if the parameters can be transferred from one type of plant to another. In this case, we suggest to create a scenario with multiple plants in the simulation.

Furthermore, the original version of the presented algorithm required deterministic data, i.e., evaluating one design point multiple time would generate the same response. Thus, it could not be applied directly to non-deterministic, so-called noisy, data. Noise, e.g. caused by measurement errors, is a common problem in real-world settings. In principle, there are two approaches to cope with noisy output data:

1. *Repeated measurements.* This approach, estimates y at $\mathbf{x}$ by the mean of repeated measurements and this value is fed to the Kriging model.
2. *Nugget Effect and Reinterpolation.* In this approach, the Kriging model is able to handle the raw, noisy data. The noise can simply be handled in the Kriging model by fitting it with a so-called nugget effect. However, this will lead to non-zero uncertainty estimates at already tested locations in the search space. This will decrease the effectiveness of using expected improvement. A solution through reinterpolation is offered by [7].

Finally, the surrogate-modeling method employed here may profit from computationally cheaper information that can be gained from a coarse variant of the objective function. In detail, a parameterization of the mill model may be tested with less data. This results into faster evaluation times, but yields a less accurate estimation of the quality. This cheaper information can be used in tandem with the more accurate, expensive data to train a better surrogate-model. To that end, Co-Kriging [6] can be employed. Co-Kriging allows to exploit correlation between coarse and fine target functions. This may result into a much improved surrogate-model of the hot-mill, without the requirement of additional expensive evaluations.

References

1. Barton, R.R., Meckesheimer, M.: Metamodel-based simulation optimization. In: S.G. Henderson, B.L. Nelson (eds.) Simulation, *Handbooks in Operations Research and Management Science*, vol. 13, pp. 535 – 574. Elsevier (2006). DOI http://dx.doi.org/10.1016/S0927-0507(06)13018-2. URL http://www.sciencedirect.com/science/article/pii/S0927050706130182
2. Bartz-Beielstein, T.: Experimental analysis of evolution strategies—overview and comprehensive introduction. Interner Bericht des Sonderforschungsbereichs 531 Computational Intelligence CI–157/03, Universität Dortmund, Germany (2003)
3. Bartz-Beielstein, T., Lasarczyk, C., Preuß, M.: Sequential parameter optimization. In: B. McKay, et al. (eds.) Proceedings 2005 Congress on Evolutionary Computation (CEC'05), Edinburgh, Scotland, vol. 1, pp. 773–780. IEEE Press, Piscataway NJ (2005)
4. Bartz-Beielstein, T., Parsopoulos, K.E., Vrahatis, M.N.: Design and analysis of optimization algorithms using computational statistics. Applied Numerical Analysis and Computational Mathematics (ANACM) **1**(2), 413–433 (2004)
5. Bartz-Beielstein, T., Zaefferer, M.: A gentle introduction to sequential parameter optimization. Tech. Rep. TR 01/2012, CIplus (2012)
6. Forrester, A., Sóbester, A., Keane, A.: Multi-fidelity optimization via surrogate modelling. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science **463**(2088), 3251–3269 (2007). DOI 10.1098/rspa.2007.1900
7. Forrester, A., Sobester, A., Keane, A.: Engineering Design via Surrogate Modelling. Wiley (2008)

8. Hajduk, M., Zidek, M., Elfmark, J., Kopec, S.: Derivation of mean values of inherent deformation resistance in hot rolling of tonnage steel. Hutnicke Listy **27**(8), 567 (1972)

9. Hajduk, M., et al.: Effect of improper selection of the rpm of vertical and horizontal drives on balanced rolling force distribution in a universal rolling mill. Hutnicke Listy **27**(8), 259 (1972)

10. Hensel, A., Spittel, T.: Kraft- und Arbeitsbedarf bildsamer Formgebungsverfahren. Verlag Grundstoffindustrie (1978)

11. Hernandez, C., Medina, S., Ruiz, J.: Modelling austenite flow curves in low alloy and microalloyed steels. Acta Materialia **44**(1), 155 – 163 (1996). DOI http://dx.doi.org/10.1016/1359-6454(95)00153-4. URL http://www.sciencedirect.com/science/article/pii/1359645495001534

12. Hinkfoth, R.: Massivumformung. Wissenschaftsverlag, Aachen (2003)

13. Jin, Y.: A comprehensive survey of fitness approximation in evolutionary computation. Soft Computing **9**(1), 3–12 (2005)

14. Jones, D., Schonlau, M., Welch, W.: Efficient global optimization of expensive black-box functions. Journal of Global Optimization **13**, 455–492 (1998)

15. Jones, D.R.: A taxonomy of global optimization methods based on response surfaces. J. of Global Optimization **21**, 345–383 (2001). DOI http://dx.doi.org/10.1023/A:1012771025575. URL http://dx.doi.org/10.1023/A:1012771025575

16. Kleijnen, J.P.C.: Design and analysis of simulation experiments. Springer, New York NY (2008)

17. Lambiase, F.: Optimization of shape rolling sequences by integrated artificial intelligent techniques. The International Journal of Advanced Manufacturing . . . **68**(1-4), 443–452 (2013)

18. Mancini, E., Campana, F., Sasso, M., Newaz, G.: Effects of cold rolling process variables on final surface quality of stainless steel thin strip. The International Journal of Advanced Manufacturing . . . **61**(1-4), 63–72 (2012)

19. Mandal, S., Rakesh, V., Sivaprasad, S., Venugopal, S., Kasiviswanathan, K.V.: Constitutive equations to predict high temperature flow stress in a Ti-modified austenitic stainless steel. Materials Science and Engineering (2009)

20. Nelder, J., Mead, R.: A simplex method for function minimization. Computer Journal **7**, 308–313 (1965)

21. Powell, M.: A review of algorithms for nonlinear equations and unconstrained optimization. In: Proceedings ICIAM, pp. 220–232 (1988)

22. Powell, M.: A direct search optimization method that models the objective and constraint functions by linear interpolation. Tech. Rep. DAMTP 1992/NA5, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England (1992)

23. Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P.: Design and analysis of computer experiments. Statistical Science **4**(4), 409–435 (1989)

24. Sheu, J.J.: Simulation and optimization of the cold roll-forming process. In: AIP, pp. 452–457. AIP, Melville, NY (2004)

25. Sun, J., Du, F., Li, X.: FEM Simulation of the Roll Deformation of Six-high CVC Mill in Cold Strip Rolling. In: 2008 International Workshop on Modelling, Simulation and Optimization (WMSO), pp. 412–415. IEEE (2008)

26. Tselikov, A., Nikitin, G., Rokotyan, S.: The Theory of Lengthwise Rolling. Mir Publishers (1981)

27. Weber, K.: Grundlagen des Bandwalzens. VEB Deutscher Verlag fuer Grundstoffindustrie, Leipzig (1973)

28. Y.C. Lin Ming-Song Chen, J.Z.: Prediction of 42crmo steel flow stress at high temperature and strain rate. Meachanics Research Communications (2008)