Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Online Model Adaptation in Cold Rolling for Improvement of Thickness Precision ⋆

Matthias Wehr * **David Stenger** *,⋆⋆ **Sven Schätzler** **
Ralf Beyer * **Dirk Abel** * **Gerhard Hirt** **

* *Institute of Automatic Control, RWTH Aachen University, 52062 Aachen, Germany (Tel: +49-241-80-28015; e-mail: m.wehr@irt.rwth-aachen.de)*
** *Institute of Metal Forming, RWTH Aachen University, 52062 Aachen, Germany (e-mail: sven.schaetzler@ibf.rwth-aachen.de)*

Abstract: Cold rolling is a process that finishes the production of flat steel and must therefore guarantee high strip precision. However, the strip thickness produced in the roll gap cannot be measured directly which makes its observation in the roll gap challenging. In this paper, the model of both the mill frame as well as the cold rolled strip are optimized online using measured process data. A Recursive Least Squares parameter estimator is used to determine mill modulus and offset of the roll stand, while the rolling model of the steel strip is adapted using Gaussian Process Regression. The adapted models are then used in a model based controller which adjusts the roll gap accordingly. Experimental results show that the precision of the models is enhanced using online measurements. As a result the desired strip thickness is achieved despite initial model uncertainties.

Keywords: Cold rolling, Model adaptation, Gaussian Process Regression, Recursive Least Squares Estimation, Self-calibration, Real-time control

1. INTRODUCTION

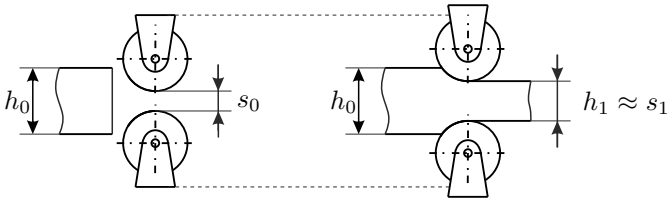
Cold rolling is a metal forming process where a work piece is primarily reduced in thickness. This is done by passing the sheet through a pair of rolls which rotate in opposite directions with a manipulated roll gap size of s_0. When pulled through the rolls the strip is reduced from the incoming strip thickness h_0 to the outcoming strip thickness h_1 with the roll force F. The most important quality measure is the outcoming strip thickness which is desired to match the reference thickness $h_{1,\mathrm{des}}$ with as little deviation as possible. Both the mill frame and the strip are influenced by the occurring force as can be seen in Fig. 1. Thus, not only the strip is reduced in thickness, but also the mill frame is extended by the force F. This leads to the loaded roll gap size s_1 with a mill stretch

$$\Delta s = s_1 - s_0. \tag{1}$$

As a result, the manipulated roll gap s_0 does not match the outcoming thickness h_1. Instead, the loaded roll gap approximately matches the outcoming strip thickness. Neglecting elastic recovery effects of the strip, it holds

$$h_1 \approx s_1, \tag{2}$$

where $h_0 > h_1 \approx s_1 > s_0$ with the roll force $F > 0$. The incoming strip thickness h_0 can be measured in advance



(a) Neutral gap ($F = 0$) (b) Loaded gap ($F > 0$)

Fig. 1. Characteristic behavior of the mill frame (derived from Kopp and Wiegels (1999, in German))

while the roll force F and the roll gap s_0 can be acquired in the roll gap. Nevertheless, measuring h_1 in the roll stand is not possible. Thus, it can only be monitored model wise based on a process model using F. Fig. 2 depicts the strip and the position of the thickness gauges which measure $h_{0,\mathrm{m}}$ and $h_{1,\mathrm{m}}$. Distances between the gauges and the roll

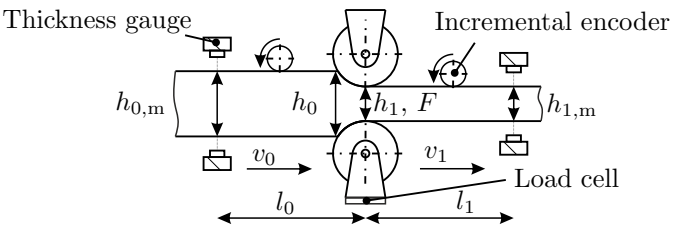


Fig. 2. Rolling mill with thickness gauges and incremental encoders to track measured values along l_0 and l_1, modified from Wehr et al. (2018)

gap are denoted as l_0 and l_1 while strip speeds are denoted by v_0 and v_1, respectively. A comprehensive introduction to flat rolling is given by e.g. Lenard (2014).

⋆ The authors appreciate the financial support of the German Research Foundation (DFG) received within the research project no. 267239860 "Enhancement of width and thickness tolerances of metallic strips by using a piezo-electric control system for roll gap adjustment of a tandem mill."
⋆⋆David Stenger is a non-funded associate of the research training group GRK 1856 of the German Research Foundation.

Knowing the strip thickness h_1 in the roll gap is, however, desirable for any kind of thickness control. Thus, two models have to be developed which determine h_1 based on the measured quantities. The first effect to be modeled is the appearing mill frame extension force F_M which depends on the difference $s_1 - s_0$ and a set of mill frame parameters $\boldsymbol{\Theta}_\mathrm{M}$ including the mill modulus (stiffness):

$$F_\mathrm{M} = f(s_1 - s_0, \boldsymbol{\Theta}_\mathrm{M}). \tag{3}$$

The second effect is the roll force F_S which is modeled as a function of h_0 and h_1 with several parameters $\boldsymbol{\Theta}_\mathrm{S}$:

$$F_\mathrm{S} = f(h_0, h_1, \boldsymbol{\Theta}_\mathrm{S}). \tag{4}$$

The parameter vector $\boldsymbol{\Theta}_\mathrm{S}$ includes coefficients such as friction, yield stress, strip tension, roll diameter, and others. More details are given by Alexander (1972) and Stockert et al. (2017). While rolling both forces have the same quantity $F = F_\mathrm{M} = F_\mathrm{S}$ which is explained in further detail in section 3.

In prior work experiments on a rolling mill for slit strips were conducted. A Model Predictive Controller (MPC) was implemented for roll gap control. It receives the reference trajectory for s_0 according to h_0 and the desired thickness $h_{1,\mathrm{des}}$ from the above mentioned nominal process models (Wehr et al., 2018; Stockert et al., 2018). In such a way a feedforward strip thickness controller was established. Although high accuracy was achieved in the experiments a high calibration and modeling effort was involved. On the one hand the roll gap s_0 needs to be calibrated accurately. On the other hand the force needed for strip thickness reduction needs to be modeled as precisely as possible, too. However, due to inevitable model uncertainties, calibration errors and changing process conditions, the desired strip thickness could not be reached only using models which were derived offline.

In literature, several adaptive modeling approaches for both model types have been presented. The mill modulus is modeled linearly by e.g. Kugi et al. (2000), Lenard (2014), and Kopp and Wiegels (1999, in German). Also, experiments conducted by Wehr et al. (2019) show a linear behavior of the mill frame of the used cold rolling mill beyond a certain force. Hence, linear parametric modeling is favorable. This is different for the cold rolling model. Models which are based on the elementary theory of plasticity and which are considered to deliver high precision results (Bland and Ford, 1948; Alexander, 1972) carry several material and process parameters which have nonlinear impact. A verification was done by e.g. Stockert et al. (2017). Also, there are approaches to adapt those models online. Randall and Stephens (1997) for example use parameter estimation by Extended Kalman Filter (EKF) based on a hot rolling model in the first roll stand of a section mill. The results are passed to all subsequent roll stands. For use in an EKF those models need to be simplified and observable. Pires et al. (2009) perform parametric identification of friction and yield stress based on the model provided by Bland and Ford (1948) using Nelder-Mead method. Ohta and Washikita (2006) use a Recursive Least Squares (RLS) algorithm for online adaptation of a simplified process model for a downstream roll stand. These approaches bear the chance of increasing process knowledge by means of white box modeling, yet they need to be simplified or complex optimization strategies have to be adopted which may be prone to local minima.

Another way of modeling is using data-driven approaches from the machine learning domain such as artificial neural networks (ANN). They do not rely on a physically motivated model structure and therefore can incorporate unanticipated effects. The ANN is parametrized by its number of layers and neurons per layer which must be defined in advance. Depending on the number of neurons, much data is needed to train the ANN. Thus, data can be generated by a highly accurate physical model (Zárate and Bittencout, 2008) or finite element method (Shahani et al., 2009) to train an ANN. Clearly, if measured data is available the models can be trained and used in the next roll pass as done by Zheng et al. (2018). In adaptive control ANN have further been used where the model is trained offline by the above mentioned method while thickness controller parameters are tuned using online data (Mousavi Takami et al., 2010). By simplification of the learning process i.e. using incremental learning together with additional offline learning it was demonstrated, that online learning is possible with ANN (Son et al., 2005). A drawback using standard ANN is that the favorable number of neurons and layers is not known for a certain model in advance. Furthermore, the model uncertainty of sample points not covered by the data can neither be quantified nor can any extrapolation guarantee to converge to a technical useful value.

Assuming that the desired thickness $h_{1,\mathrm{des}}$ is not attained initially because of model uncertainties and incorrect calibration, ANN cannot guarantee to find a correct extrapolation to the desired point. Instead, a method with good extrapolation properties is required which can relate the values of $h_{1,\mathrm{des}}$ and the actually measured values of h_1 in the proximity of the desired value. Furthermore, in completely unknown areas of the process it should converge to the precalculated nominal model. Both criteria are fulfilled by the Gaussian Process Regression (GPR). GPR is sometimes also known as *kriging* and is a non-parametric probabilistic regression method widely used in various domains. In comparison to other black box modeling approaches such as neural networks, GPR may result in similar precision for prediction but requires far less training data (Kamath et al., 2018). In literature, GPR is famously known in geosciences, based on the approach of Matheron (1971). In mining applications it is used to extrapolate the spread of resources for nearby but unknown regions (Daya Sagar et al., 2018). In recent years GPR has been widely used to map nonlinear relationships between model inputs and outputs (Rasmussen and Williams, 2006). Moreover, there are applications in the chemical industry where GPR is used for soft sensors (Grbic et al., 2012; Fan et al., 2014; Zhang et al., 2015). These approaches enable process monitoring even if only sensors with poor sample rates are used. In order to use only recent data points, moving window strategies are implemented for data selection. In context of control the mean model identified by GPR is used e.g. by Hewing et al. (2018) in a nonlinear MPC which aims for lap time optimization of a racing car.

This paper aims at the improvement of process quality i.e. strip thickness precision by enhancing the underly-

ing models involved in rolling using data available in the process for online learning. The roll gap size of the mill cannot be calibrated precisely in the magnitude of the desired tolerances. For this purpose the roll stand characteristic curve (3) is adapted in mill modulus and calibration offset using RLS estimation. Moreover, a non-parametric cold rolling model is adapted to compensate model uncertainties in (4) which arise from an initially calculated rolling model. Using GPR, the precalculated cold rolling model is expected to be improved even if the initial values do not hit the desired operating point exactly. Here, the measurement of h_1 is especially important not only because it is the controlled variable, but also because it is the common variable present in both models.

Hence, the presented work extends the MPC presented by Wehr et al. (2018) by an adaptation mechanism for the predetermined models to consider current process conditions and thus to close the loop for thickness control.

The paper organizes as follows: Section 2 briefly summarizes the most important fundamentals and aspects of the used algorithms. In section 3 it is shown how the process is modeled along with its nominal conditions. Afterwards, section 4 explains how the above mentioned algorithms are utilized to adapt the process models online. Section 5 shows the controller architecture as well as the distribution of calculation to several cores of a real-time system. Section 6 presents the results of the conducted experiments. Finally, section 7 concludes the results and gives an outlook on further work.

2. FUNDAMENTALS

2.1 Recursive Least Squares optimization

If a model to be estimated is known in its structure and carrying linear parameters $\mathbf{\Theta}$, RLS estimation can be used to find these parameters. The estimated parameter vector $\hat{\mathbf{\Theta}}$ is chosen such that the quadratic deviation between model output y and measurement value y_m is minimized. In order to be able to integrate new values without recalculating the entire set of data points a recursive formulation e.g. given in (5) can be used for current time k using information from the previous time $k-1$:

$$\hat{\mathbf{\Theta}}_\mathrm{k} = \hat{\mathbf{\Theta}}_{\mathrm{k}-1} + \mathbf{g}_\mathrm{k} \cdot \left(y_\mathrm{m} - \mathbf{m}_\mathrm{k}^\mathrm{T} \hat{\mathbf{\Theta}}_{\mathrm{k}-1} \right), \tag{5}$$

where $\mathbf{g}_\mathrm{k}$ is the correction factor and $\mathbf{m}_\mathrm{k}$ is the measurement vector which bears the structure of the system. The correction factor is updated as follows:

$$\mathbf{g}_\mathrm{k} = \frac{\mathbf{P}_{\mathrm{k}-1}\mathbf{m}_\mathrm{k}}{\rho + \mathbf{m}_\mathrm{k}^\mathrm{T}\mathbf{P}_{\mathrm{k}-1}\mathbf{m}_\mathrm{k}}, \tag{6}$$

where $\mathbf{P}$ is the covariance matrix of the estimated parameters and ρ is the forgetting factor. Moreover, $\mathbf{P}$ is updated in every time step k as well using

$$\mathbf{P}_\mathrm{k} = \frac{1}{\rho}\left(\mathbf{I} - \mathbf{g}_\mathrm{k}\mathbf{m}_\mathrm{k}^\mathrm{T}\right)\mathbf{P}_{\mathrm{k}-1}, \tag{7}$$

where $\mathbf{I}$ is the identity matrix. In such a way, computational effort can be kept minimal since large matrix inversions are avoided and the parameter vector $\hat{\mathbf{\Theta}}$ is updated in every time step. Further reading of system identification with RLS is provided e.g. by Isermann (2005, p. 303)

2.2 Gaussian Process Regression

Hereafter, a brief introduction to GPR is given. For a detailed introduction the reader is referred to Rasmussen and Williams (2006, chap. 2). From now on we consider n observations (measurements) with inputs $\mathbf{X} = [x_1, \ldots, x_n]$ and targets $\mathbf{Y} = [y_1, \ldots, y_n]$. They are combined in the data set $\mathbf{D} = [\mathbf{X}, \mathbf{Y}]$. In GPR, it is assumed that a measurement $y(x) = f(x) + \epsilon$ with input $x \in \mathbb{R}^D$ is corrupted by additive noise ϵ, where $f(x)$ represents the true unknown function value at location x_i or x_j. GPR is a non parametric probabilistic regression model. It does not rely on assuming a fixed set of basis functions for $f(x)$ as for example in polynomial regression or artificial neural networks. Instead it is assumed that any finite set of function evaluations at different locations $\{f_1, \ldots, f_n, \} = \{f(x_1), \ldots, f(x_n)\}$ is distributed according to a multivariate normal distribution. This relation can be written as a Gaussian Process:

$$(f_\mathrm{i}) \sim GP(m(x_\mathrm{i}, \theta_\mathrm{m}), k(x_\mathrm{i}, x_\mathrm{j}, \theta_\mathrm{k})). \tag{8}$$

The GP is fully characterized by an a-priori mean $m(x, \theta_\mathrm{m}) = \mathbf{E}[f(x)]$ with hyper parameters θ_m and a covariance function $k(x_\mathrm{i}, x_\mathrm{j}, \theta_\mathrm{k}) = \mathrm{cov}(f(x_\mathrm{i}), f(x_\mathrm{j}))$ with hyper parameters θ_k. Applying bayes rule to Equation (8) allows to calculate the conditional normal distribution of f at an unknown location as a function of previously obtained function values. Since we are in a setting where sensor measurements are noisy, we consider the case where observations are corrupted by homoscedastic gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma_\mathrm{n}^2)$ with variance σ_n^2. The true function value $f \sim \mathcal{N}(\bar{f}, \sigma_\mathrm{f}^2)$ is considered to be a predictive normal distribution with expected value $\bar{f}$ and its variance σ_f^2. Then, the predictive distribution of the observed value is

$$y|f \sim \mathcal{N}(f, \sigma_\mathrm{n}^2) = \mathcal{N}(\bar{f}, \sigma_\mathrm{f}^2 + \sigma_\mathrm{n}^2) \tag{9}$$

That way, the predictive distribution with mean $\bar{y}(x_*) = \bar{f}(x_*)$ and accumulated variance $\sigma_\mathrm{y}^2(x_*) = \sigma_\mathrm{f}^2(x_*) + \sigma_\mathrm{n}^2$ for a measurement at a given location x_* can be written as:

$$\begin{aligned} y(x_*) \sim \mathcal{N}(&\bar{f}(x_*|\mathbf{D}, \theta_\mathrm{m}, \theta_\mathrm{k}, \sigma_\mathrm{n}^2), \\ &\sigma_\mathrm{y}^2(x_*|\mathbf{D}, \theta_\mathrm{m}, \theta_\mathrm{k}, \sigma_\mathrm{n}^2)). \end{aligned} \tag{10}$$

By making suitable choices for $m(x, \theta_\mathrm{m})$ and $k(x_\mathrm{i}, x_\mathrm{j}, \theta_\mathrm{k})$ as well as for the observation model, we can encode prior knowledge about the particular system. Without any measurements, the predicted mean is exactly identical to the a-priori mean function $m(x, \theta_\mathrm{m})$. Additionally, if no training data is available in the proximity of a query point x_* where the model is evaluated, the prediction will tend towards the very same a-priori mean. A so called kernel function (covariance) $k(x_\mathrm{i}, x_\mathrm{j}, \theta_\mathrm{k})$ is used to determine how well the responses of two points $f(x_\mathrm{i})$ and $f(x_\mathrm{j})$ correlate given their locations x_i and x_j. In this work an isometric squared exponential kernel

$$k(x_\mathrm{i}, x_\mathrm{j}, \theta_\mathrm{k}) = \sigma_0^2 \mathrm{exp}\left[-\frac{1}{2} \sum_{d=1}^{D} \left(\frac{x_{\mathrm{i,d}} - x_{\mathrm{j,d}}}{l_\mathrm{s}} \right)^2 \right] \tag{11}$$

with hyper-parameters $\theta_k = \{\sigma_0^2, l_\mathrm{s}\}$ is chosen. Correlation between two locations decreases with increasing distance. The quantitative impact is determined by the length scale l_s and the kernel scaling factor σ_0^2. Omitting the details, equations for predictive mean and predictive variance at multiple query points $\mathbf{X}_*$ are given in (12) and (13). The predictive mean $\bar{f}_*$ is given by

$$\bar{f}_* = \bar{f}(\mathbf{X}_*) = m(\mathbf{X}_*) + \mathbf{K}(\mathbf{X}_*, \mathbf{X}) \cdot \mathbf{K}_\mathrm{Y}^{-1} \cdot (\mathbf{Y} - m(\mathbf{X})) \quad (12)$$

while the estimated variance $\mathbf{\Sigma}_{f_*}$ is calculated by

$$\mathbf{\Sigma}_{f_*} = \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*) - \mathbf{K}(\mathbf{X}_*, \mathbf{X}) \cdot \mathbf{K}_\mathrm{Y}^{-1} \cdot \mathbf{K}(\mathbf{X}, \mathbf{X}_*). \quad (13)$$

Both equations include m query points $\mathbf{X}_*$ as well as the measurement points $\mathbf{X}$. Furthermore, they require the inversion of $\mathbf{K}_\mathrm{Y}$ given by

$$\mathbf{K}_\mathrm{Y} = \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_\mathrm{n}^2 \cdot \mathbf{I}. \quad (14)$$

The covariance matrix $\mathbf{K}$ can be calculated for any combination of $\mathbf{K}(\mathbf{X}, \mathbf{X})$, $\mathbf{K}(\mathbf{X}_*, \mathbf{X})$, and $\mathbf{K}(\mathbf{X}, \mathbf{X}_*)$ following the scheme

$$\mathbf{K}(\mathbf{X}, \mathbf{X}_*) = \begin{bmatrix} k(x_1, x_{*1}, \theta_\mathrm{k}) & \dots & k(x_1, x_{*\mathrm{m}}, \theta_\mathrm{k}) \\ \vdots & \ddots & \vdots \\ k(x_\mathrm{n}, x_{*1}, \theta_\mathrm{k}) & \dots & k(x_\mathrm{n}, x_{*\mathrm{m}}, \theta_\mathrm{k}) \end{bmatrix}. \quad (15)$$

3. PROCESS MODELS

3.1 Roll stand model

Most roll stands have characteristics similar to a progressive spring. The initial part is nonlinear, relatively soft and barely generates any force, while at larger stretch the roll stand shows linear behavior (Lenard, 2014, p. 106), (Wehr et al., 2019). Therefore, the roll stand can be modeled linearly by a parameter vector $\hat{\mathbf{\Theta}}_\mathrm{M} = [a_1 \ a_0]^\mathrm{T}$ with the mill modulus a_1 and an additional offset a_0 beyond the limiting force F_lim as shown in Fig. 3. Beyond F_lim it holds:

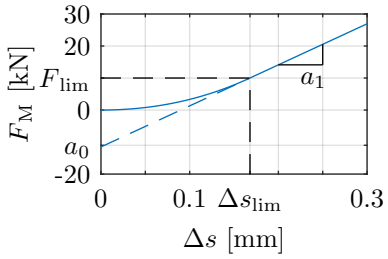


Fig. 3. Nominal roll stand characteristic curve

$$F_\mathrm{M} = a_1 \Delta s + a_0. \quad (16)$$

Applying (2) yields

$$F_\mathrm{M} = a_1 (h_1 - s_0) + a_0. \quad (17)$$

Parameterization can be done offline by pressing the work rolls against each other without an inserted strip such that $h_1 = s_1 = 0$. By variation of s_0 the resulting F_M can be measured. In such a way $\hat{\mathbf{\Theta}}_\mathrm{M}$ is initially determined to

$$\hat{\mathbf{\Theta}}_\mathrm{M} = \left[128.23\,\mathrm{kN\,mm^{-1}} \ \ -11.52\,\mathrm{kN}\right]. \quad (18)$$

Furthermore, the limiting force is set to $F_\mathrm{lim} = 10\,\mathrm{kN}$.

3.2 Cold rolling model

As mentioned in section 1, the roll force F_S depends on h_0, h_1 and several parameters $\mathbf{\Theta}_\mathrm{S}$ which influence the force in a nonlinear way. Moreover, due to roll flattening as well as a varying contact length between the rolls and the strip, iterative calculation is necessary. Therefore, the roll force model is pre-calculated offline using the model developed by Alexander (1972) with the nominal parameters $\mathbf{\Theta}_\mathrm{S} = \{\mu, t_0, t_1, \nu, E, n, B, Y_0, r, b\}$ given in Tab. 1. The resulting map is depicted in Fig. 4 (blue). It can be used to calculate

Table 1. Cold rolling model parameters

Variable	Value	Description
μ	0.11	Friction
t_0	100 N	Incoming strip tension
t_1	100 N	Outcoming strip tension
ν	0.3	Poisson's ratio
E	210 GPa	Young's modulus
n	0.165	Swift coefficient
B	193.4	Swift constant
Y_0	270 MPa	Initial strain
r	60 mm	Rigid roll radius
b	18.9 mm	Strip width

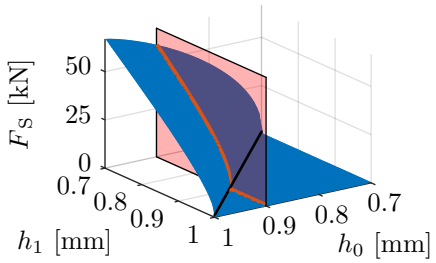


Fig. 4. Nominal cold rolling force map

the force at particular pairs of thicknesses $\{h_0, h_1\}$. An exemplary force needed to reduce the strip thickness from $h_0 = 0.9$ to any h_1 is highlighted in red. The force is zero beyond $h_1 = 0.9\,\mathrm{mm}$ and increases nonlinearly from $h_1 = 0.9\,\mathrm{mm}$ to $h_1 = 0.7\,\mathrm{mm}$. The advantage of the map is that a force can be evaluated rapidly for any pair $\{h_0, h_1\}$. The disadvantage, clearly, is that changes in parameters lead to a recalculation of the whole map.

3.3 Combined model

To combine both models, it is needed to extract the curve at a specific h_0 from the rolling map and find the intersection with the roll stand line (see Fig. 5). If h_0 changes, the curve to be extracted from the map changes as well. The position of the roll stand line, again, depends on the roll gap size s_0. The resulting h_1 and F are obtained at the intersection of both curves. Smaller values of s_0, i.e. a smaller roll gap, will move the line to the left resulting in a smaller h_1 and a larger F. A larger s_0 value will result in the opposite. Thus, if h_0 and s_0 as well as the roll gap models are known precisely, h_1 and F can be calculated.

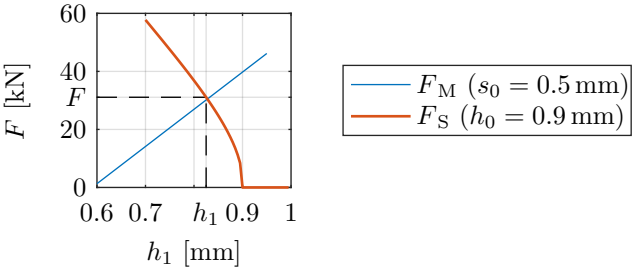


Fig. 5. Nominal roll force, strip thickness diagram

4. SELF-CALIBRATION AND MODEL ADAPTATION

The process models described above are nominal models of the process. They are crucially important for process quality when used for strip thickness control. Hence, they

have to be adapted during the process to account for wrong calibration and model perturbations. Fortunately, both models can be adapted independently from each other. It is assumed that the roll stand characteristics as well as strip parameters such as yield stress and friction in the roll gap change much slower than the measurement transport delay l_1 of h_1. Thus, the model is considered valid even if a delayed h_1 is used for model adaptation (see Fig. 2).

4.1 Roll stand estimation

Due to the progressive characteristics of the roll stand starting with zero slope, it is difficult to determine the offset of the mill modulus by means of a micrometer and the initial point can easily be missed (see. Fig. 3). Thus, the true neutral roll gap s_0 is considered unknown. Furthermore, the mill modulus may change after a change of the roller pack or after grinding of the rolls. Thus, s_0 has to be divided into a known and an unknown part:

$$s_0 = s_0' + s_{0,\mathrm{p}}, \tag{19}$$

where s_0' is the manipulated roll gap which can be sensed and $s_{0,\mathrm{p}}$ is the calibration offset. It is important to emphasize, that only s_0' can be actuated. Assuming that rolling only takes place in the linear part of the roll stand curve it can be modeled by means of (17). Inserting (19) yields:

$$F = a_1 \cdot (h_1 - (s_0' + s_{0,\mathrm{p}})) + a_0 \tag{20}$$

Since neither the calibration offset, nor the linear offset are known, only an accumulated quantity a_0^* can be estimated.

$$F = a_1 \cdot (h_1 - s_0') + \underbrace{a_0 - a_1 \cdot s_{0,\mathrm{p}}}_{a_0^*} \tag{21}$$

In such a way a_0^* includes both the linear offset as well as any calibration offset regarding the roll gap. Thus, the true neutral roll gap remains unknown. However, for a given parameterization of $\{a_1, a_0^*\}$ the function $f : s_0' \to F$ is known. Thus, the generated force F can be calculated depending only on the sensed roll gap size s_0'. The roll stand model is put into the shape shown in (22) in order to be fit for Recursive Least Squares estimation:

$$\begin{aligned} y_\mathrm{m} &= F, \\ \mathbf{m}_\mathrm{k}^\mathrm{T} &= [h_1 - s_0' \;\; 1], \\ \hat{\mathbf{\Theta}}_\mathrm{M} &= [a_1 \;\; a_0^*]^\mathrm{T}. \end{aligned} \tag{22}$$

An initial guess $\hat{\mathbf{\Theta}}_\mathrm{M}$ can be obtained by means of a roughly calibrated line e.g. (18). Since the RLS is valid only in the linear part beyond F_lim, it must be assured that enough force is generated initially such that valid measurements can be acquired. Since a static line is estimated, no dynamics have to be considered. Hence, there is no need to include consecutive time steps. This allows selection of valid data whenever they occur.

The covariance of the parameters can be set according to the initial trust in the parameters. From (21) it can be seen, that a_1 can be determined in advance since it depends only on $\Delta s'$. Hence, the covariance of a_1 is chosen very small while that of a_0^* is chosen higher:

$$\mathbf{P}_0 = \begin{bmatrix} 0.01 & 0 \\ 0 & 100 \end{bmatrix}. \tag{23}$$

In such a way, primarily a_0^* will be adapted. Since it is likely that the mill modulus is stable over a very long time, the forgetting factor is chosen negligible low: $\rho = 1 - 10^{-8}$. In order to avoid discontinuities in the roll gap controller caused by sudden changes of the parameters, they are passed through a rate limiter to the control loop.

4.2 Cold rolling model adaptation

The GPR methodology presented in section 2 is used to adapt the rolling model. It considers n measurements with inputs $\mathbf{X}$ where $x_\mathrm{i} = [h_{0,\mathrm{i}}, h_{1,\mathrm{i}}]$ and targets $\mathbf{Y}$ where $y_\mathrm{i} = F$. These measurements are used to predict the roll force across the domain $h_0 = [0.98\,\mathrm{mm}, 1.02\,\mathrm{mm}]$ and $h_1 = [0.88\,\mathrm{mm}, 0.98\,\mathrm{mm}]$. The nominal force F_S obtained from Alexander's model is chosen as the prior mean $m(x, \theta_\mathrm{m})$ of the GPR model, where $\theta_\mathrm{m} = \mathbf{\Theta}_\mathrm{S}$. As described in section 2 GPR ensures that if no measurements are available in a certain region of the $h_0 - h_1$ surface the predictive mean tends towards the nominal model F_S. Therefore, the roll force is only adapted locally in an area where measurements are available. As described in Section 2 an isometric squared exponential kernel is used. A squared exponential kernel with different length scales for each dimension (SEARD) is not necessary, because variations in h_0 and h_1 are expected to influence the prediction on the same length scale with similar magnitudes. For the hyper parameters of the GPR, $l_\mathrm{s} = 0.028\,\mathrm{mm}, \sigma_n = 1.648\,\mathrm{kN}$ and $\sigma_0 = 1\,\mathrm{kN}$ are chosen using domain knowledge i.e. the expected mutual influence of the measured values is adjusted based on the authors' experience. [1]

As can be seen from (12), the standard case GPR results in a cubic time complexity problem regarding the number of measurements due to the inversion of $\mathbf{K}_\mathrm{Y}$. This prohibits recalculating the predictive mean $\bar{f}_*$ and variance $\mathbf{\Sigma}_{f_*}$ of the GPR model in each time step for a growing amount of measurements. Therefore, in order to limit the computational cost, only 250 measurements are considered for model adaptation. In order to minimize the information loss when using only the latest measurements, a grid-based heuristic is used to select suitable points for adaptation. The selection heuristic aims at favoring recent observations $\mathbf{D}$ over past ones while preserving a distribution within the $h_0 - h_1$ domain. For this purpose, the $h_0 - h_1$ domain is divided into a grid of 55 equally sized segments in which the observations are assigned. In order to keep the number of considered observations constant the oldest observation in the segment with the most existing entries is discarded when a new point is added to the grid.

The prediction is done for a map of a fixed range in the $h_0 - h_1$ domain instead of calculating a single point in every time step. It is spanned by means of a 41×41 grid and covers the range of h_0 from 0.98 mm to 1.02 mm as well as h_1 from 0.88 mm to 0.98 mm. This results in the matrix of multiple query points $\mathbf{X}_* \in \mathbb{R}^{1681 \times 2}$ which contains every grid point x_* of the map $\bar{f}_*$. It is calculated at discrete time intervals of $T_\mathrm{s} = 2\,\mathrm{sec}$ and transferred to the controller afterwards. Whenever the controller queries a particular prediction of the force $\hat{F}_\mathrm{S}$ on the map, it is calculated by interpolation in $\mathbf{X}_*$.

[1] Preliminary experiments have shown that choosing the hyper parameters by maximizing the likelihood resulted in unsatisfactory results. This can be explained by the fact that the data set was not representative.

5. REAL-TIME IMPLEMENTATION

5.1 Synchronization of measurement data

For all equations stated above it is assumed that all values are valid and present for calculation. In fact, they are recorded simultaneously at different locations in the rolling mill. Therefore, they need to be spatially synchronized. This is done using several speed-independent buffers (Wehr et al., 2018). The value of the gauge for the incoming strip thickness $h_{0,\mathrm{m}}$ is tracked to the roll gap with speed v_0 along distance l_0. There, s'_0 and F are collected and tracked to gauge $h_{1,\mathrm{m}}$ with strip speed v_1 along distance l_1 where $h_{1,\mathrm{m}}$ is collected (recap Fig. 2). All values are transported another 0.5 m beyond the position of $h_{1,\mathrm{m}}$. In such a way spatial filtering as shown by Wehr et al. (2018) can be done for all values. Here, a cutoff period of $D_\mathrm{c} = 0.151\,\mathrm{m}$ is chosen. Since the filter removes periods smaller than D_c, these inputs can neither be sensed nor do they disturb the model adaptation in terms of noise. Once all values are synchronized and filtered, they are processed by the presented optimization algorithms.

5.2 Controller topology

The experiments are carried out on a dSPACE DS 1006 real-time machine. Online adaptation through machine learning algorithms such as GPR may require a lot of computational resources. Therefore, they are calculated using a dedicated core on the real-time machine. Fig. 6 shows the topology of three dedicated CPUs with all measured quantities (Meas). Communication between the processors

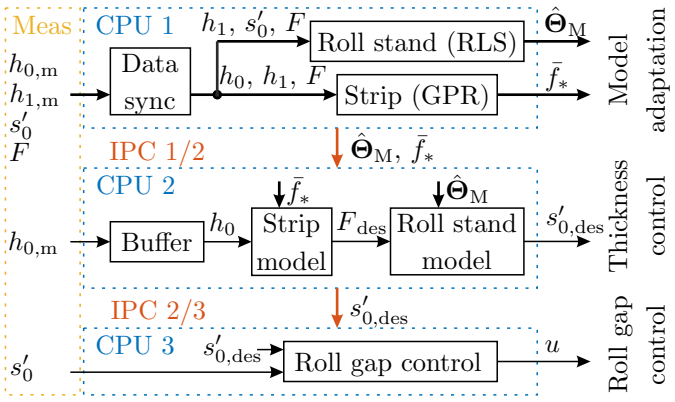


Fig. 6. Thickness control block diagram

is realized using inter processor communication (IPC). CPU 1 performs the above mentioned model adaptation and provides the estimated parameters $\hat{\boldsymbol{\Theta}}_\mathrm{M}$ and the cold rolling map $\bar{f}_*$ to the thickness controller running on CPU 2. The sample time of the GPR is set to $T_\mathrm{s} = 2\,\mathrm{sec}$. A major drawback is, that despite splitting the applications to several cores, the calculation of (13) exceeds the memory on the machine due to large matrix multiplications. Thus, no information on the process variance could be generated within scope of this work. CPU 2 determines a desired roll gap according to a buffered incoming strip thickness trajectory using the adapted models. This is done at $T_\mathrm{s} = 0.001\,\mathrm{sec}$. CPU 3 receives the desired roll gap size $s'_{0,\mathrm{des}}$ and performs closed loop tracking of s'_0 with manipulated variable u at $T_\mathrm{s} = 0.001\,\mathrm{sec}$.

6. EXPERIMENTAL RESULTS

In the experiment the nominal thickness of the incoming strip is $h_0 = 1\,\mathrm{mm}$. The desired thickness is $h_1 = 0.95\,\mathrm{mm}$. In order to guarantee valid values for the estimation algorithms and to ensure safe operation of the rolling mill within the process constraints from the very beginning, the estimated parameters should be in the proximity of the true parameters. This makes conscientious modeling of the process indispensable yet leaves room for uncertainties, model errors, and calibration offsets which can be corrected by the online adaptation.

6.1 Adaptation of the mill housing parameters

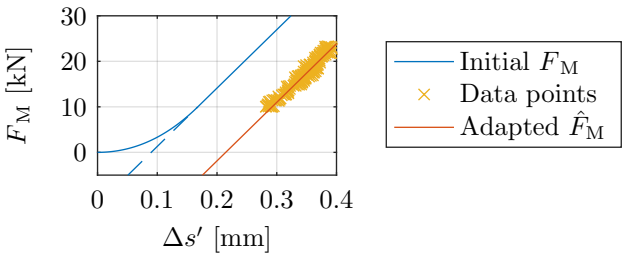


Fig. 7. Adaptation of the mill housing line

Fig. 7 depicts the result of the adapted mill characteristic line. The blue dashed line is initially determined in (18). The red line is found by the RLS algorithm yielding

$$\hat{\boldsymbol{\Theta}}_\mathrm{M} = \begin{bmatrix} 128.60\,\mathrm{kN\,mm^{-1}} & -27.59\,\mathrm{kN} \end{bmatrix}. \tag{24}$$

By comparing the initial guess with the RLS solution a significant error e in the offset a_0^* can be noted. The difference between the initial and final values yields $e = \begin{bmatrix} 0.37\,\mathrm{kN\,mm^{-1}} & -16.07\,\mathrm{kN} \end{bmatrix}$. Thus the assumption of the inaccurate calibration of a_0^* holds true while at the same time it can be compensated by the estimation.

6.2 Adaptation of the cold rolling model

Fig. 8 depicts the result of the GPR model adaptation. The blue plane shows the initial model while the red one shows the latest adapted model. The yellow dots show all available data points. Among those, the white data points are chosen by the heuristic selector for data set **D**. Initially, all points are located in the neighborhood of $(1.01\,\mathrm{mm}, 0.98\,\mathrm{mm})$ due to wrong model calibrations. After the adaptation the points are accumulated in the desired region $(1.01\,\mathrm{mm}, 0.95\,\mathrm{mm})$. Tab. 2 shows a numeric comparison of three exemplary points. It can be seen that the model adaptation compensated a model error of $-1.97\,\mathrm{kN}$ at the desired strip thickness pair $(1.01\,\mathrm{mm}, 0.95\,\mathrm{mm})$. In addition, the nominal rolling model almost covered the measurements at the pair $(1.01\,\mathrm{mm}, 0.98\,\mathrm{mm})$ that was initially hit. The third pair shown is a point which is never reached by measurement values $(1.005\,\mathrm{mm}, 0.96\,\mathrm{mm})$. However, the predicted force lies in the proximity of the other points which suggests that the extrapolated value is valid. Fig. 9 depicts the section view of the models. The red lines show the nominal (dashed) as well as the adapted cold rolling models for $h_0 = 1.01\,\mathrm{mm}$. The blue lines show the roll stand line for the latest $\hat{\boldsymbol{\Theta}}_\mathrm{M}$ at different s'_0. It can be seen that the initial

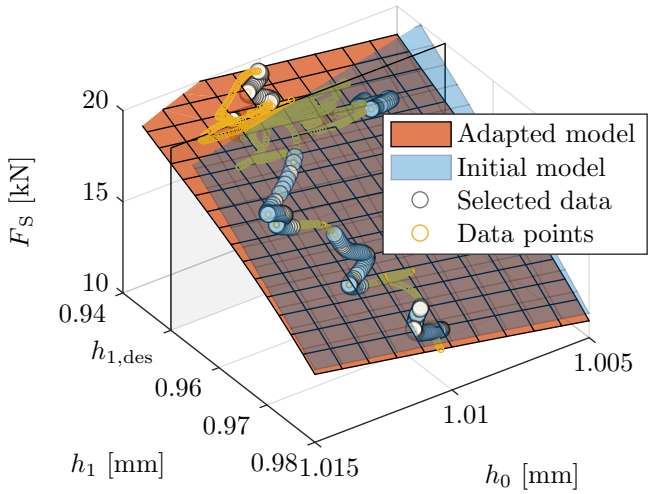


Fig. 8. Cold rolling characteristic map adaptation

(dashed) roll gap size s'_0 is chosen too large which leads to a small thickness reduction (as will be shown in Fig. 10). Once $\hat{\boldsymbol{\Theta}}_M$ is found the roll gap size is decreased leading to a larger thickness reduction as depicted by the blue non-dashed line. In order to achieve $h_1 = 0.95\,\text{mm}$, a roll gap

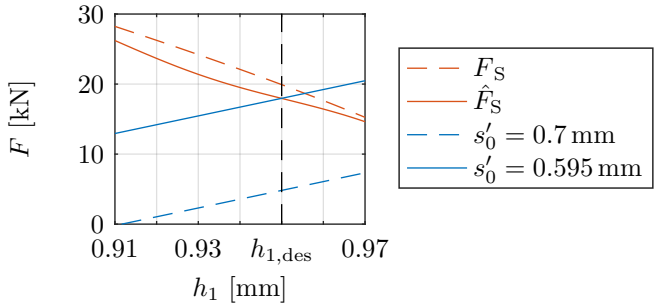


Fig. 9. Roll force / strip thickness section view

size of $s'_0 = 0.595\,\text{mm}$ is necessary. Fig. 9 shows a drawback of the method. While the absolute force needed is tracked up to a small deviation its derivative may deviate from the nominal model significantly. This is due to the fact that $\bar{f}_*$ converges to the measurements where they are available and to the initial model where no measurements are available. Therefore, the resulting map $\bar{f}_*$ needs to be treated carefully when being derived for the use in e.g. a state observer.

6.3 Strip thickness enhancement

Fig. 10 shows the first 5 m of the synchronized strip. Initially both models calculate wrong values for F and s'_0 due to inaccurate calibration. Thus, only little force is applied and strip thickness reduction is minimal. Once valid values are processed by the synchronization the adaptation algorithms start working. The algorithm detects that the

Table 2. Cold rolling model comparison

Pair (h_0, h_1)	F_S	$\hat{F}_S$	e
(1.01 mm, 0.95 mm)	19.92 kN	17.95 kN	−1.97 kN
(1.01 mm, 0.98 mm)	12.69 kN	12.35 kN	−0.34 kN
(1.005 mm, 0.96 mm)	16.49 kN	15.21 kN	−1.28 kN

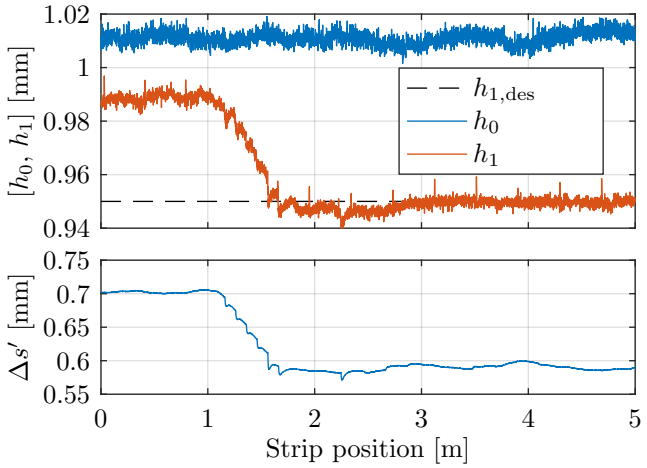


Fig. 10. Strip thickness result

applied roll gap is too far open which impacts the applied force. At about one meter the rate-limited adaptation starts working. As a result the roll gap is closed slowly which yields towards the desired strip thickness reduction. At this point the cold rolling map determines rather too much force. This is caused by the effect that the data points in the neighborhood of (1.01 mm, 0.98 mm) are very close to the model such that there is no need to further reduce the level of $\bar{f}_*$ around the desired thickness. This leads to an undershoot of the strip thickness. At 3 m the GPR adapts the cold rolling map in the desired region, too, and the strip converges to the reference thickness. After the models have converged at about 4 m it can be seen that small deviations in h_0 are compensated by changing s'_0. This leads to a more constant profile in h_1.

7. CONCLUSION AND OUTLOOK

In this paper, a parametric online identification of the mill housing characteristics and a non-parametric identification of the rolling model were presented. Both aim at an enhancement of the combined process model which could be used for strip thickness control and thus strip quality enhancement. It was shown that an adjustment of the roll gap based on valid process models leads to the desired outcoming strip thickness. Although the true roll gap remains unknown the adaptation of the mill housing line could be used to determine an actuation which delivers the required force for tracking the reference thickness. Furthermore, the adaptation of the cold rolling model was used in order to determine the required force for the particular strip. In such a way the control loop could be closed accounting for long-term process behavior. This led to an improvement of the previous work (Wehr et al., 2018). Due to convergence to the nominal model in regions without data points the model behavior is predictable. This favors application in a productive environment. During the experiments it turned out that the result strongly depends on the data selection. For application in production a sophisticated data selection method has to be used. In order to account for short-term disturbances and roll eccentricities, a roll gap observer has to be developed which uses the adapted models for estimation of the outcoming strip thickness based on the measured force. While not being able to determine the process variance is not an issue within scope of this work, it can be used to determine the validity of the

adapted model in particular regions. This could be used by e.g. further control loops. Thus, future work needs to focus on efficient calculation of the process variance, too.

REFERENCES

Alexander, J.M. (1972). On the theory of rolling. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 326(1567), 535–563. doi:10.1098/rspa.1972.0025.

Bland, D.R. and Ford, H. (1948). The calculation of roll force and torque in cold strip rolling with tensions. *Proceedings of the Institution of Mechanical Engineers*, 159(1), 144–163. doi: 10.1243/PIME_PROC_1948_159_015_02.

Daya Sagar, B.S., Cheng, Q., and Agterberg, F. (2018). *Handbook of Mathematical Geosciences*. Springer International Publishing. doi:10.1007/978-3-319-78999-6.

Fan, M., Ge, Z., and Song, Z. (2014). Adaptive gaussian mixture model-based relevant sample selection for jitl soft sensor development. *Industrial & Engineering Chemistry Research*, 53(51), 19979–19986. doi: 10.1021/ie5029864.

Grbic, R., Sliskovic, D., and Kadlec, P. (2012). Adaptive soft sensor for online prediction based on moving window gaussian process regression. In *11th International Conference on Machine Learning and Applications*, 428–433. IEEE. doi:10.1109/ICMLA.2012.160.

Hewing, L., Liniger, A., and Zeilinger, M.N. (2018). Cautious nmpc with gaussian process dynamics for autonomous miniature race cars. In *2018 European Control Conference (ECC)*, 1341–1348. IEEE. doi: 10.23919/ECC.2018.8550162.

Isermann, R. (2005). *Mechatronic Systems - Fundamentals*. Springer Science & Business Media, London.

Kamath, A., Vargas-Hernández, R.A., Krems, R.V., Carrington, T., and Manzhos, S. (2018). Neural networks vs gaussian process regression for representing potential energy surfaces: A comparative study of fit quality and vibrational spectrum accuracy. *The Journal of chemical physics*, 148(24), 241702. doi:10.1063/1.5003074.

Kopp, R. and Wiegels, H. (1999, in German). *Einführung in die Umformtechnik*. Verlag Mainz, Aachen, 2. edition.

Kugi, A., Haas, W., Schlacher, K., Aistleitner, K., Frank, H.M., and Rigler, G.W. (2000). Active compensation of roll eccentricity in rolling mills. *IEEE Transactions on Industry Applications*, 36(2), 625–632. doi: 10.1109/28.833781.

Lenard, J.G. (2014). *Primer on Flat Rolling*. Elsevier, 2nd edition.

Matheron, G. (1971). *The Theory of Regionalized Variables and Its Applications*. Centre de Morphologie Mathématique Fontainebleau: Les cahiers du Centre de Morphologie Mathématique de Fontainebleau. École national supérieure des mines.

Mousavi Takami, K., Mahmoudi, J., and Dahlquist, E. (2010). Adaptive control of cold rolling system in electrical strips production system with online-offline predictors. *The International Journal of Advanced Manufacturing Technology*, 50(9-12), 917–930. doi: 10.1007/s00170-010-2585-7.

Ohta, T. and Washikita, Y. (2006). Adaptive control for the head-end strip gauge using recursive least squares at hot strip mill. In *2006 IEEE Conference on Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control*, 1831–1836. IEEE. doi:10.1109/CACSD-CCA-ISIC.2006.4776919.

Pires, C.T.d.Á., Ferreira, H.C., and Sales, R.M. (2009). Adaptation for tandem cold mill models. *Journal of Materials Processing Technology*, 209(7), 3592–3596. doi:10.1016/j.jmatprotec.2008.08.020.

Randall, A. and Stephens, R.I. (1997). On-line adaptive control in the hot rolling of steel. *IEEE Proceedings - Control Theory and Applications*, 144(1), 15–24. doi: 10.1049/ip-cta:19970990.

Rasmussen, C.E. and Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*. MIT Press.

Shahani, A.R., Setayeshi, S., Nodamaie, S.A., Asadi, M.A., and Rezaie, S. (2009). Prediction of influence parameters on the hot rolling process using finite element method and neural network. *Journal of Materials Processing Technology*, 209(4), 1920–1935. doi: 10.1016/j.jmatprotec.2008.04.055.

Son, J.S., Lee, D.M., Kim, I.S., and Choi, S.G. (2005). A study on on-line learning neural network for prediction for rolling force in hot-rolling mill. *Journal of Materials Processing Technology*, 164-165, 1612–1617. doi: 10.1016/j.jmatprotec.2005.01.009.

Stockert, S., Wehr, M., Lohmar, J., Abel, D., and Hirt, G. (2017). Assessment of flat rolling theories for the use in a model-based controller for high-precision rolling applications. In *20th International ESAFORM Conference on Material Forming*. Dublin, Ireland. doi: 10.1063/1.5008218.

Stockert, S., Wehr, M., Lohmar, J., Hirt, G., and Abel, D. (2018). Improving the thickness accuracy of cold rolled narrow strip by piezoelectric roll gap control at high rolling speed. *CIRP Annals*, 67(1), 313–316. doi: 10.1016/j.cirp.2018.04.107.

Wehr, M., Stockert, S., Abell, D., and Hirt, G. (2018). Model predictive roll gap control in cold rolling with piezoelectric actuators. In *2018 IEEE Conference on Control Technology and Applications (CCTA)*, 1377–1382. doi:10.1109/CCTA.2018.8511333.

Wehr, M., Stockert, S., Ionescu, C., Abel, D., and Hirt, G. (2019). Sliding mode control of piezoelectric stack actuators for roll gap adjustment in a cold rolling mill. In *2019 International Conference on Advanced Intelligent Mechatronics (AIM)*, 1207–1214. IEEE/ASME. doi: 10.1109/AIM.2019.8868531.

Zárate, L.E. and Bittencout, F.R. (2008). Representation and control of the cold rolling process through artificial neural networks via sensitivity factors. *Journal of Materials Processing Technology*, 197(1-3), 344–362. doi: 10.1016/j.jmatprotec.2007.06.063.

Zhang, W., Li, Y., Xiong, W., and Xu, B. (2015). Adaptive soft sensor for online prediction based on enhanced moving window gpr. In *2015 International Conference on Control, Automation and Information Sciences (ICCAIS)*, 291–296. IEEE. doi: 10.1109/ICCAIS.2015.7338679.

Zheng, G., Ge, L.H., Shi, Y.Q., Li, Y., and Yang, Z. (2018). Dynamic rolling force prediction of reversible cold rolling mill based on bp neural network with improved pso. In *2018 Chinese Automation Congress (CAC)*, 2710–2714. IEEE. doi:10.1109/CAC.2018.8623139.