Material matters: predicting the core hardness variance in industrialized case hardening of 18CrNi8

Vorhersage der Kernhärtenvarianz von industriell einsatzgehärtetem 18CrNi8

Y. Lingelbach[1], T. Waldenmaier[2], L. Hagymasi[1], R. Mikut[3], V. Schulze[4]

To explain the variance in core hardness of 18CrNi8 nozzle bodies after industrial heat treatment, several data sources, including steel melt composition, sensor process data, and measurement errors, of five years are aggregated. In order to predict hardness variations caused by alloy composition, traditional physical models by Maynier are compared with data-driven machine learning models, which show no advantage due to low data variability. Neither method can fully explain the visible drifts, which are better tracked by an alternative (i.e., filter model) that uses past measurements. Machine learning on features from heat treatment is not successful in predicting hardness change, presumably because the process is too stable. Finally, a large part of the variance is caused by the HV 1 measurement error.

Keywords: Case hardening / Alloy elements / Prediction / Machine learning / Data driven / Variance / Hardness / Industrial process

Um die Varianz der Kernhärte von Düsenkörpern aus 18CrNi8 nach einer industriellen Wärmebehandlung zu erklären, werden mehrere Datenquellen, darunter die Werkstoffzusammensetzung der Stahlschmelze, die Prozessdaten der Sensoren sowie die Messfehler, der vergangenen fünf Jahren kombiniert und analysiert. Zur Vorhersage von Härteänderungen, die durch die Werkstoffzusammensetzung verursacht werden, werden traditionelle physikalische Modelle von Maynier datengetriebenen maschinellen Lernmodellen gegenübergestellt. Keine der beiden Methoden kann die sichtbaren Drifts vollständig erklären, die durch ein alternatives Filtermodell, das vergangene Messungen verwendet, besser erfasst werden. Maschinelles Lernen auf Basis von Merkmalen aus der Wärmebehandlung ist bei der Vorhersage der Härteänderung nicht erfolgreich; vermutlich aufgrund der hohen Prozessstabilität und der damit einhergehenden zu geringen Varianz. Schließlich

[1] Robert Bosch GmbH, Powertrain Solutions Division - Manufacturing Technology Management, Stuttgart, Germany

[2] Robert Bosch GmbH, Corporate Sector Research and Advance Engineering -Materials- and Process Engineering Metals, Renningen, Germany

[3] Karlsruhe Institute of Technology, Institute for Automation and Applied Informatics (IAI), Hermann-von-Helmholtz-Platz 1, Eggenstein-Leopoldshafen, Deutschland

[4] Karlsruhe Institute of Technology, Institute for Applied Materials (IAM), Karlsruhe, Deutschland

Corresponding author: Y. Lingelbach, Robert Bosch GmbH, Powertrain Solutions Division - Manufacturing Technology Management, Wernerstraße 51, 70469, Stuttgart, Germany,
E-Mail: Yannick.Lingelbach.YL@gmail.com

wird ein großer Teil der insgesamt zu beobachtenden Varianz durch die Streuung der HV 1 Messung verursacht.

1 Introduction

Reducing cost and quality variances in established heat treatment processes becomes more difficult the more optimization measures such as increased batch size and reduced process time have already been implemented. To quantitatively assess the factors' relative contribution to the variance in core hardness of case hardened nozzle bodies, data from two production stages and their quality assessment are analyzed: steel production, case hardening, and hardness measurement, *Figure 1*. Over 90 % of the hardness values measured fall within an interval of 40 HV 1, suggesting that the individual effects may be difficult to access if their accumulation is limited to such a narrow range.

To the best of the authors' knowledge, the concatenation and relative influence of the individual hardness effects of such an industrialized case hardening process have not yet been investigated in the available literature. Moreover, no real industry-like application data were used but mainly data from the literature or from smaller experiments, which are elaborated on in the following.

Hardness can be predicted by application of data driven methods such as supervised learning based on chemical composition and tensile properties [1, 2]. Additional knowledge about the austenitization conditions, at least for low alloy steels, makes predictions even better [2, 3]. The ability to predict hardness from chemical composition and additional parameters (e. g., austenitization and tempering temperatures as well as duration) can be learned by artificial neural networks when trained on data available from the literature (i. e., tables in books) or simulated data [4–8]. On this basis, machine learning methods may also predict the martensite/bainite/austenite start temperature or volume fraction of bainite [9–12]. Naturally, sufficiently large changes in chemical composition also predicts mechanical properties such as Charpy toughness, yield strength and contact fatigue life [13–15].
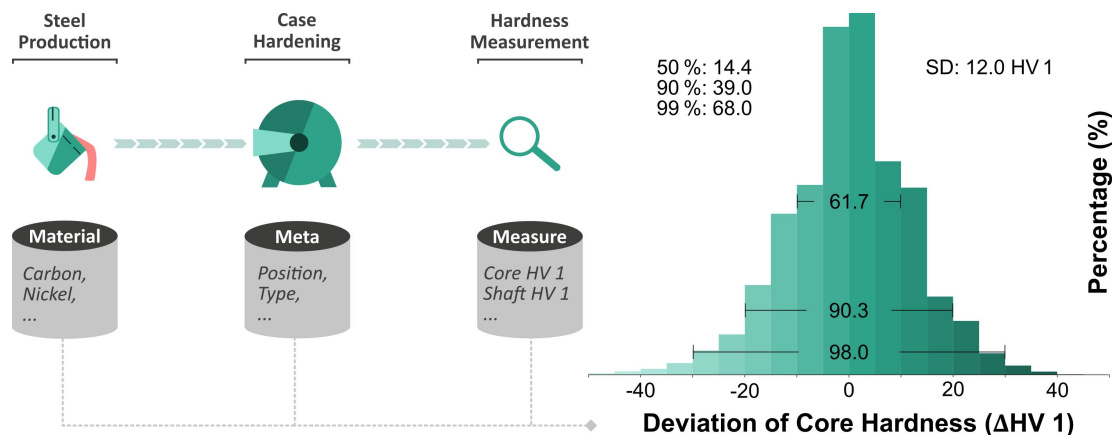

Figure 1. Schematic representation of the process chain. Data was collected from steel production (material composition of the steal melt), to heat treatment (meta data about the furnace, component type, etc.) and hardness measurements. The resulting core hardness distribution with a standard deviation (SD) of 12 HV 1 stems from ~ 7000 HV 1 measurements over 5 years.

Bild 1. Schematische Darstellung der Prozesskette. Es wurden Daten von der Stahlherstellung (Materialzusammensetzung der Stahlschmelze) über die Wärmebehandlung (Metadaten: Ofen, Bauteiltyp, usw.) bis hin zu den Härtemessungen erfasst. Die sich daraus ergebende Kernhärteverteilung mit einer Standardabweichung (SD) von 12 HV 1 stammt von ~ 7000 HV 1-Messungen über 5 Jahre.

In the present paper, the authors investigate whether these models are also of predictive value for the very small variations experienced in high quality steels that must meet strict industry quality criteria and are monitored over long periods of time in industrial production. Due to the lack of available data, publications on hardness predictions from industrialized heat treatment processes are rare. Supervised machine learning has been applied to bainitization with some success using process parameters as inputs to predict the hardness. However, these approaches missed to take material composition into account and failed to use a chronological train test split [16]. Industrial data has not yet been used for the hardness prediction from case hardening process, but the heat treatment and its hardening results have been thoroughly simulated [17]. Examination of the reproducibility of hardness measurements shows that considerable variance is to be expected due to calibration deviation of the hardness testing machines, the influence of the specimen preparation as well as diamond abrasion [18, 19].

The state of the art for predicting the hardness of heat-treated steels is capable of accurately describing individual self-contained stages of the production process but has not been applied to industrial data or to the entire process chain. The present work fills this gap and investigates the practicality of hardness prediction in an industrial setting, taking into account all important factors in the process chain and their relative significance.

2 Materials and methods

2.1 Process chain and data

The injection nozzle bodies are manufactured from 18CrNi8 (1.5920) case hardening steel supplied by Stahl Judenburg GmbH to a certified degree of purity, *Table 1*.

While nozzle body blank production (not shown in Figure 1) is a single piece flow, the subsequent heat treatment consists of 7200 pieces in one batch (6-layer rag), where layers may contain different types of nozzle bodies (i.e., slight variation of geometry). The heat treatment process consists of three stages: 1. case hardening, that is, carburization and quenching in one of three vacuum furnaces, 2. cooling in one of two deep freezers, and 3. passing through one of three tempering furnaces, all controlled by the same system Prosys by DEMIG. Batch paths are variable, that is, after deep freezer #1 or #2 the batch may go through tempering furnace #1, #2, or #3. For quality assurance, a defined number of nozzle bodies are tested after heat treatment. From each batch, two nozzle bodies with alternating positions are sampled (i.e., components from position c1 and c4 for batch i, components from position f3 and f6 for batch i + 1), *Figure 2*. This information is important, as an individual nozzle body's closeness to the door, center or quenching funnel is expected to influence its heat treatment. Each sampled nozzle body is cut lengthwise, embedded, and polished. An HV 1 indentation is used to evaluate the hardness in Vickers (HV) at various positions and distances from the surface. This paper focuses on two measurement positions: the core 3 mm away from the surface (Core) and the shaft 0.7 mm away (Shaft 0.7).

Table 1. Upper and lower limits of Robert Bosch GmbH order specification of the weight fractions [wt.–%] of each alloying element.

Tabelle 1. Obere und untere Grenzwerte der Robert Bosch GmbH Bestellangaben für die Gewichtsanteile [wt.–%] der einzelnen Legierungselemente.

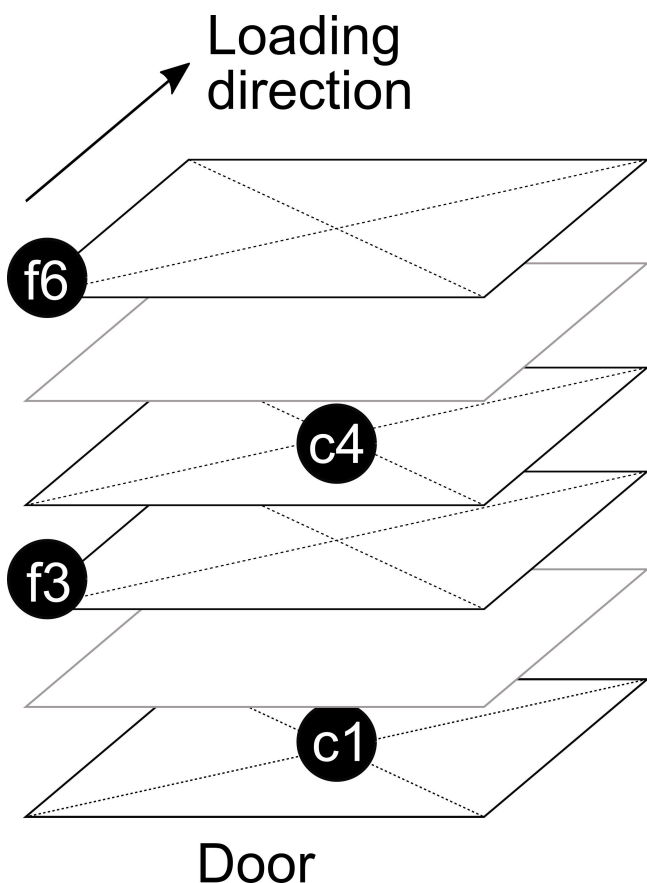|	Cr	Ni	C	Mn	Si	Al	Mo	P	S
Upper limit	2.100	2.150	0.220	0.640	0.300	0.040	0.150	0.035	0.035
Lower limit	1.700	1.750	0.130	0.360	–	0.015	–	–	–

Loading direction

Door

Figure 2. Nozzle body batch in form of a 6-layer rag filled with 7200 pieces, where c1, c4 (center), and f3, f6 (front) mark the sampling points of the test specimens.

Bild 2. Düsenkörpercharge in Form eines 6-lagigen Gestells, das mit 7200 Teilen gefüllt ist, wobei c1, c4 (center) und f3, f6 (front) die Entnahmestellen der Prüfkörper markieren.

Process as well as quality data is stored in an SQL database for reasons of traceability, quality assurance, and long-term behavior evaluation.

For the present analysis, data was collected from the beginning of 2015 to the beginning of 2020, including the following data sets: chemical composition of 283 steel melts of 18CrNi8, process data from 4480 case hardened nozzle body batches, and the hardness of 6935 HV 1 indentions per measurement position. The number of indentions is smaller than twice the number of batches, because one batch may contain different component types and not all types of specimens can be included in the analysis due to their different material composition (e.g., some nozzle body types are made from X40CrMoV5-1) and geometry. Each component type has a unique designation (e.g., A1, A2, A3,

B1, etc.), with the letter indicating the broader family-often associated with a particular customer-and the number indicating a particular characteristic of that family. Geometry can differ to varying degrees both between and within component families. The measurement error was estimated from 4 HV 1 indentions on each of 68 nozzle bodies.

2.2 Calculate and learn hardness from material composition

2.2.1 Calculation with known models

The influence of chemical composition and austenitization conditions on hardness for low alloy steels has been presented by Maynier and colleagues at the Creusot-Loire Laboratory [2, 3] and was reviewed by [20]. Depending on the volume fractions of the microstructure constituents (i.e., %FP = ferrite and perlite, %B = bainite, %M = martensite) and their respective hardness, the total hardness of the steel is calculated as a weighted sum over the mixed structures, *Equation 1* [20].

$$HV = \frac{\%FP \cdot HV_{F-P} + \%B \cdot HV_B + \%M \cdot HV_M}{100} \quad (1)$$

Due to the good-natured hardening properties of 18CrNi8 and high quenching pressure, even the microstructure in the core of the nozzle body is predominantly martensitic, with some bainitic content, after heat treatment. Therefore, the following equations can be used to estimate the hardness at the core, *Equations 2 to 5* [20]:

$$HV_M^{=127} + 949C + 27Si + \\ 11Mn + 16Cr + \\ 8Ni + 21 \cdot \log V_R \quad (2)$$

$$\log V_{M100\%} = 9.81 - (4.62C + 1.05Mn + \\ 0.5Cr + 0.66Mo + 0.54Ni + 0.00183 \cdot P_a) \quad (3)$$

$$\log V_{M90\%} = 8.76 - (4.04C + 0.96Mn + \\ 0.58Cr + 0.97Mo + 0.49Ni + \\ 0.001 \cdot P_a) \quad (4)$$

$$V_B = -H323 + 185C + 330Si + 153Mn + 144Cr + 191Mo + 65Ni +$$

$$\log V_R \cdot \begin{pmatrix} 89 + 53C - 55Ni - 22Mn - \\ 20Cr - 33Mo - 10Ni \end{pmatrix} \qquad (5)$$

Where HV_M and HV_B is the hardness calculated from the relative mass contents of the alloying elements carbon (C), silicon (Si), manganese (Mn), chromium (Cr), nickel (Ni), and molybdenum (Mo) and V_R [°Ch⁻¹] the cooling rate for martensite at 700 °C, with for example $R = M90$ % indicating 90 % martensite and 10 % other microstructures. P_a is a parameter including austenitizing time and temperature.

Equations 1 to 5 were used to estimate the mean core hardness of nozzle bodies belonging to one steel melt after heat treatment. In order to do so, two parameters need to be estimated: The volume fraction of martensite in Equation 1 and a general offset that is subtracted from Equation 1 accounting for tempering. Ferrite and perlite fractions do not occur during heat treatment. Thus, the remaining fraction is a mixture of bainite and austenite, where the hardness influence of the latter may be omitted due to its minimal share. A first method to find these parameters is via optimization by dual annealing using the python package scipy, where the loss function is the mean squared error (MSE) between the calculated hardness values per melt and the mean of the measured hardness belonging to that melt [21], *Equation 6*:

$$MSE = \frac{1}{n} \sum_{i=1}^{N} (y_i - \widehat{y}_i)^2, \qquad (6)$$

where y is the observed and $\widehat{y}$ the predicted value.

However, we report prediction results as the root of the mean squared error (RMSE) to retrieve the original unit of measurements, that is, hardness in Vickers with a load of 1 kg (HV 1). The volume fraction of martensite and the remaining microstructures can also be estimated via a simulation in Ansys® using [22].

2.2.2 Learn with linear regression

Alternatively, the coefficients for hardness prediction from the alloying elements can be determined by linear regression from the available data itself. That is, assuming that variations in an alloying element linearly increase or decrease hardness, at least for small variations around the standard chemical composition of the steel. For completeness more complex machine learning models that can map nonlinear relationships such as artificial neural networks or random forests (similar to [9, 16, 23, 24]) were also implemented by the authors, but they tend to overfit the data and therefore provide no additional benefit here. Only if the variations in material composition become sufficiently large, which does not seem to be the case with our data set, or if other factors (e. g., influences of heat treatment) are to be included (section 2.4), can it be assumed that the relationship is nonlinear. The parameters of such models are also more difficult to interpret. Moreover, extra- or and interpolation are a complicated matter for the more complex machine learning methods and their performance is strongly application dependent [25, 26].

2.3 Tracking models

To track the averaged long-term hardness fluctuations y and determine the hardness offset (i. e., the deviation of the average hardness y) caused by the specimen's batch position Δ_{Pos} and its component type Δ_{Comp}, the following model based on a first-order infinite impulse response (IIR) filter is introduced by the authors, consisting of an update step, and a prediction step, *Equations 7, 8*:

$$y_n = a \cdot y_{n-1} + (1-a) \cdot (x_n - \Delta_{Pos,n} - \Delta_{Comp,n}) \qquad (7)$$

where $a \in [0,1]$

$$\widehat{y}_{n+1} = y_n + \Delta_{Pos,\,n+1} + \Delta_{Comp,\,n+1} \qquad (8)$$

$$\widehat{y}_{n+1} = a \cdot y_{n-1} + (1 - a) \cdot \begin{pmatrix} x_n - \Delta_{Pos,n} - \\ \Delta_{Comp,n} \end{pmatrix} \quad (9)$$

$$+ \Delta_{Pos,\,n+1} + \Delta_{Comp,\,n+1}$$

where y_n [HV] is the average hardness state after production of batch n, estimated by updating its predecessor state y_{n-1} with the new measurement x_n. Because the measurement x_n is performed on of the four test specimen the offsets to the average caused by their particular batch position Δ_{Pos} and component type Δ_{Comp} must be corrected for. That is, the individual offset for position and component type are subtracted from the measurement x_n to exclude their influence on the state y. Thus, the current estimated average hardness y_n is a weighted sum between the previous average hardness state y_{n-1} and the measurement of the current batch x_n corrected by test specimen specific position and type. Given such an average state y_n, a forecast $\widehat{y}_{n+1}$ can be made about the hardness of a specific test specimen from the next batch (i.e, $n+1$) by adding the offset cause by its location $\Delta_{Pos,\,n+1}$ and component type $\Delta_{Comp,\,n+1}$ to the current average hardness y_n. The 13 parameters (i.e., a, Δ_{Pos} for four batch positions, and Δ_{Comp} for eight component types) were optimized by dual annealing. Thereby, the to-be-minimized loss function was the mean squared error between the prediction $\widehat{y}_{n+1}$ and the true measured value x_{n+1}. As a prediction $\widehat{y}_{n+1}$ is made for a particular position and component type it can only be compared to the x_{n+1} of that particular position and component. The terms y_n and x_n may not be compared as the former is an estimated average hardness of the complete batch and the latter is a measurement at a particular position in the batch.

2.4 Feature extraction and machine learning

Parameter fluctuations in the heat treatment procedure may influence the resulting hardness and are investigated by machine learning (ML) methods. In this approach, machine learning algorithms attempt to learn a relationship between input variables, also called features (e.g., austenitizing temperature), and a label (e.g., core hardness) for a subset of the data (i.e., training set) and then predict the label from the features for samples they have not yet seen (i.e., test set). Process features extracted from each heat treatment are the mean, minimum, maximum, standard deviation, and time span of the following process parameters: austenitization temperature and duration, cooling rate during quenching, deep freezing temperature and duration, as well as tempering temperature and duration. To evaluate which kind of modeling approach is best suited for the hardness prediction from these features, representative algorithms from different machine learning families for regression were trained: linear regression (linear model), multi-layer perceptron with one hidden layer (artificial neural network), random forest regressor (ensemble tree), and support vector regressor (sparse kernel method). A systematic review of machine learning algorithms and predictive models in general along with their practical application is provided in [27] while an in depth mathematical explanation can be found in [28]. For these algorithms to work accurately, their hyperparameters must be adapted to the problem (i.e., data set) at hand. This optimization was performed automatically using Bayesian search [29]. In the following the search ranges for these parameters are outlined as well as the value of those hyperparameters that differ from their default setting in the scikit-learn library [30]: linear regression (none); multi-layer perceptron (*early_stopping = True, learning_rate = adaptive, hidden_layer_sizes: Int(1,100)*); random forest regressor (*n_estimators: Int(10, 300), max_depth: Int(2,15), min_samples_split: Int(2, 30)*) and support vector regressor (*C: Real(1e–6, 1e3,'log-uniform'), epsilon: Real(1e–6, 0.99,'log-uniform')*).

The machine learning pipeline always included a robust scaler as first step and was implemented by use of the scikit-learn library [30]. Features were selected either by sequential feature forward selection or the feature importance attribute of the random forest while final hyperparameter tuning was performed manually. This is important, as too many and/or uninformative features can easily lead to overfitting, jeopardizing the generalizability of the prediction to future data sets [31].

Models were trained on 80 % of the data and tested on 20 %, once divided chronologically and once randomly. The labels were corrected by the predictions from the tracking model to ensure that

the machine learning algorithms only learn and predict hardness deviations from the current mean hardness caused by process fluctuations during heat treatment. Section 3.3 attempts to predict the hardness of nozzle bodies after heat treatment (i.e., case hardening, deep freezing, and tempering) from the measured process parameters using the machine learning methods mentioned above to explain the variance possibly caused by process variations themselves.

3 Results

3.1 Core hardness and material composition over time

The mean core hardness of the nozzle bodies for different batch positions as well as their chemical composition change over time, *Figure 3*. A diamond replacement in 2018 and a recalibration of the measurement device also significantly changed the standard deviation of the measurements from 10.6 HV 1 before 2018 to 8.9 HV 1 after the

change (standard deviation was calculated after subtracting the mean hardness of the rolling window, significance was determined by an *F*-Test with $p < .01$). For this reason, the horizontal stripes showing the measurement resolution capacity move closer together after the replacement, Figure 3a. In the following, we first examine why nozzle bodies from the two batch positions f3 and f6 are associated with different core hardness, and then consider influences related to changes in chemical composition that appear to be partially responsible for the long-term variations.

3.1.1 Influence of batch position on microstructure

Upper layer position f6 experiences stronger quenching than position f3 (i.e., mean cooling rate between 800 °C and 500 °C for f6 = 20 °C s^{-1} and f3 = 8 °C s^{-1}, as determined by a temperature uniformity survey) because it is more exposed to the quenching gas nitrogen which is first injected from the furnace ceiling. Thus, nozzle bodies on the upper layer show an increased martensite to bainite
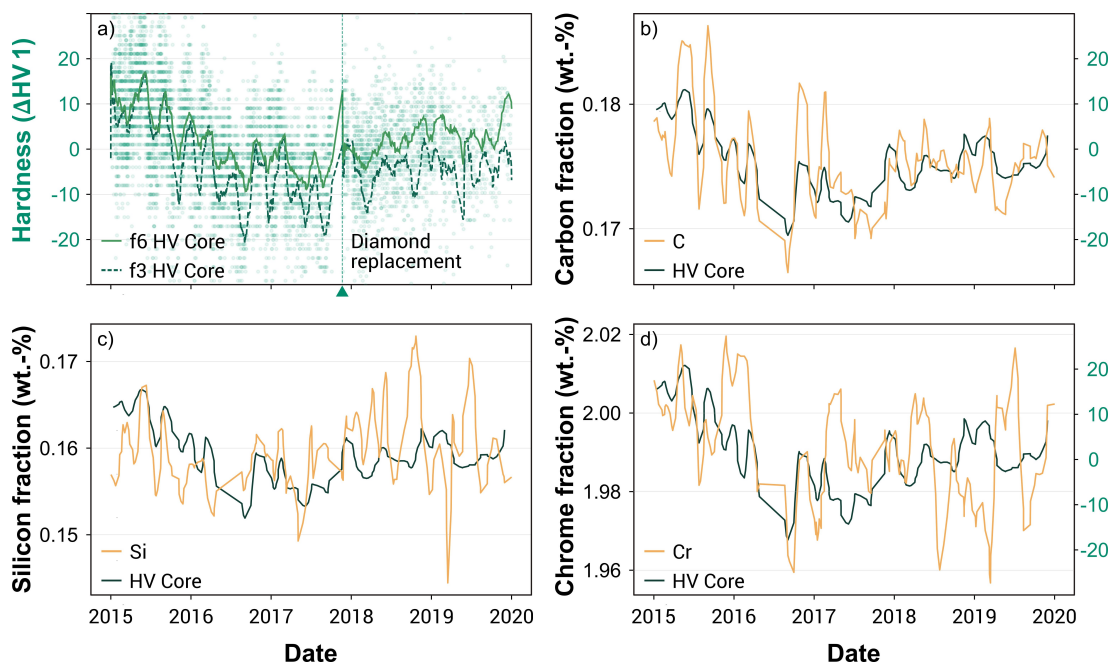

Figure 3. a) Individual measurement results as well as a rolling window with a size of 50 days of the core hardness, b), c) and d) weight fraction of carbon (C), silicon (Si) and chromium (Cr) in each steel melt as well as mean core hardness of every nozzle body belonging to that melt.

Bild 3. a) Einzelne Messergebnisse sowie der Mittelwert eines gleitenden Fensters mit einer Größe von 50 Tagen der Kernhärte, b), c) und d) Gewichtsanteil von Kohlenstoff (C), Silizium (Si) und Chrom (Cr) in jeder Stahlschmelze sowie mittlere Kernhärte jedes zu dieser Schmelze gehörenden Düsenkörpers.
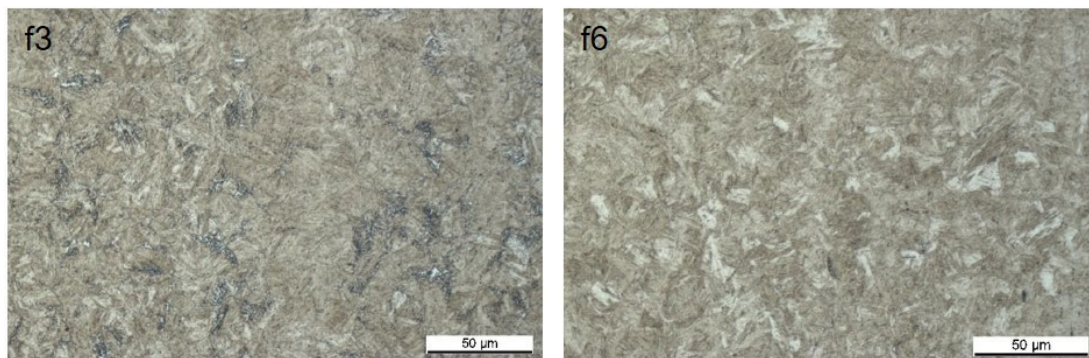
Figure 4. Microsections of nozzle bodies from batch positions f3 and f6 showing the increased bainitic (dark gray) fraction at f3 in the otherwise martensitic microstructure.

Bild 4. Schliffbilder der Düsenkörper von den Chargenpositionen f3 und f6, die den erhöhten bainitischen Anteil (dunkelgrau) bei f3 in der ansonsten martensitischen Mikrostruktur zeigen.

ratio and, therefore, greater hardness. We estimated this ratio for positions f3 and f6 using three different methods: (1) a prediction from Ansys® based on the respective cooling rates, (2) optimizing the respective parameters $\%M$ and $\%B$ in Maynier's equations for the given elemental fractions in the dataset and trying to predict the respective hardness, and (3) actual microsections from batch positions f3 and f6. Assuming that the microsection can be used as a ground truth, the Ansys® simulation overestimates the bainite content in the core of the nozzle body slightly at f3, but correctly predicts some retained austenite (not seen in the microsection), *Figure 4*, *Table 2*. Maynier's model cannot predict any retained austenite as it is not part of

Table 2. Fraction of microstructure constituents of the nozzle body core from batch positions f3 and f6 in % after heat treatment. Estimations were done by an Ansys simulation, optimization of respective parameters in Maynier's equations and actual microsections, Figure 4.

Tabelle 2. Anteil der Gefügebestandteile des Düsenkörperkerns der Chargenpositionen f3 und f6 in % nach der Wärmebehandlung. Die Abschätzungen erfolgten durch eine Ansys-Simulation, Optimierung der entsprechenden Parameter in den Maynier-Gleichungen und tatsächliche Mikroschnitte, Bild 4.

| | Martensite ($\%M$) | | Bainite ($\%B$) | | Austenite | |
	f3	f6	f3	f6	f3	f6
Ansys	73	95	26	4	1	1
Maynier	75	99	25	1	–	–
Microsection	~85	99	~15	1	–	–

its equations but comes slightly closer to the microsection estimation.

3.1.2 Influence of chemical composition on hardness

Fluctuations in chemical composition are intermittently correlated with the resulting core hardness, most notably for carbon, Figure 3b. Expectedly, carbon variation has the highest sensitivity as reflected in its comparably large coefficient in Equation 1, but it also shows the smallest fluctuation 0.02 wt.–% compared to 0.03 wt.–% for silicon. Lower carbon oscillation after 2018 likely also reduced the hardness variance for this period. To support this hypothesis, three approaches were used to predict mean core hardness from the respective chemical composition: Maynier's equations, linear regression on the whole data set (i.e., training set equals test set), and with 5-fold cross-validation (i.e., using four years as training set and predicting the fifth year).

The both calculations by Maynier's model (RMSE $=3.88$ HV 1, $R^2=0.49$) and the prediction via linear regression (RMSE $=3.77$ HV 1, $R^2=0.52$) based on the material composition, are close to the actual measured core hardness, *Figure 5*a. Linear regression with cross-validation (RMSE $=4.89$ HV 1, $R^2=0.20$) lags significantly behind the former two, indicating poor generalizability, most likely due to the fact that the diamond replacement introduces a concept drift in the data that the algorithm is not able to handle well. In
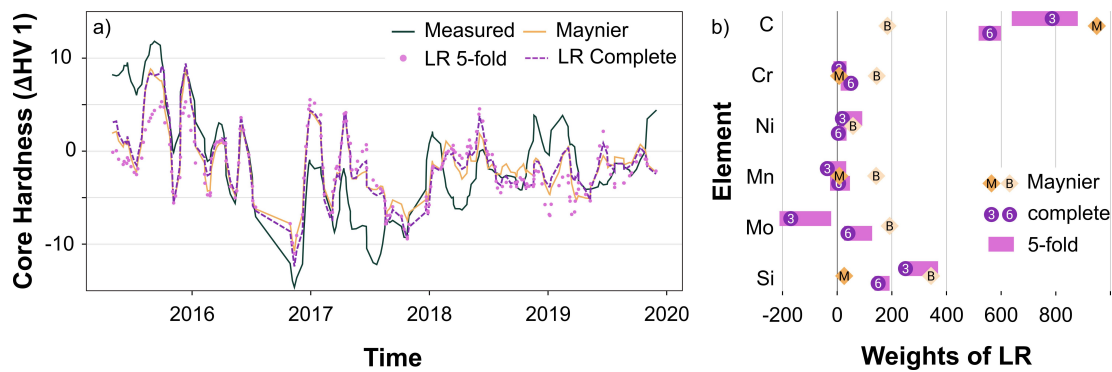
Figure 5. a) Mean core hardness of heat treated nozzle bodies per steel melt determined by measurement, Maynier's model, and linear regression, b) weights of linear regression (LR) for nozzle bodies from position f3 (lowest quenching rate) and f6 (highest quenching rate) compared to coefficients in Maynier's equations (2) for martensite (M) and (5) for bainite (B).

Bild 5. a) Mittlere Kernhärte der wärmebehandelten Düsenkörper pro Stahlschmelze, bestimmt durch Messung, Mayniers Modell und lineare Regression. b) Gewichte der linearen Regression (LR) für Düsenkörper aus Position f3 (niedrigste Abschreckrate) und f6 (höchste Abschreckrate) im Vergleich zu den Koeffizienten in Mayniers Gleichungen (2) für Martensit (M) und (5) für Bainit (B).

general, all methods are able to capture the direction of change (i. e., increasing or decreasing hardness) most of the time but frequently misjudge the amplitude. Hence, their predictions are not wrong per se, but about half of the variance (i. e., R^2 scores are ~0.5) in measured mean hardness can be explained by the material composition. When correcting for long term drifts (e. g., changes in the measurement device) by a third order polynomial, RMSEs can be reduced to 3.0 HV 1 for the first two methods. The RMSE can be slightly further reduced by 0.2 when only learning and predicting the hardness for a given batch position (e. g., position f3 and f6).

These results suggest that we may trust the models sufficiently to use their components for generalizable interpretation. Seemingly, carbon variation is responsible for most of the hardness fluctuation, as can be seen from the coefficients of the linear regression as well as from Maynier's equation, Figure 5b. Counterintuitively, the linear regression coefficients of position f3 (lowest cooling rate) are closer to Maynier's martensite coefficients than those of the position f6, suggesting that linear regression cannot learn the exact elemental influences for martensite and bainite from these data (more complex models provide no solution, as they overfit leading to worse predictions on the test set). Coefficients of the 5-fold model even switch signs for manganese and molybdenum. Linear regression also strongly overestimates the silicon influence be-

cause it is coincidentally correlated with carbon ($r = 0.56$); possibly due to a specific scrap metal the steel manufacture uses for the 18CrNi8 production. Three points become clear from this analysis: (1) Although fluctuations cannot be predicted precisely, even miniscule changes in carbon can be used to explain and are responsible for the observed core hardness oscillation. Comparing the observed changes in carbon to the allowed limits in Table 1, one needs to be aware that much more drastic fluctuations could be possible and lead to severe hardness changes. (2) Industrial datasets may not provide a good basis for extracting physical properties because minimal variation and correlated inputs hinder differentiation between individual factors. (3) A machine learning approach in this case does not yield any advantage, because the variety in the data to learn from is too small. Existing models are quasi linear and have well established coefficients and more complicated models are not necessary to explain the observations.

3.2 Forecasting hardness

Although the offset produced by batch position and component type is not particularly large, it does not seem negligible either, *Figure 6*. Merely using the mean hardness per category underestimates the influence of batch position while overshooting the component type. This is, because different compo-
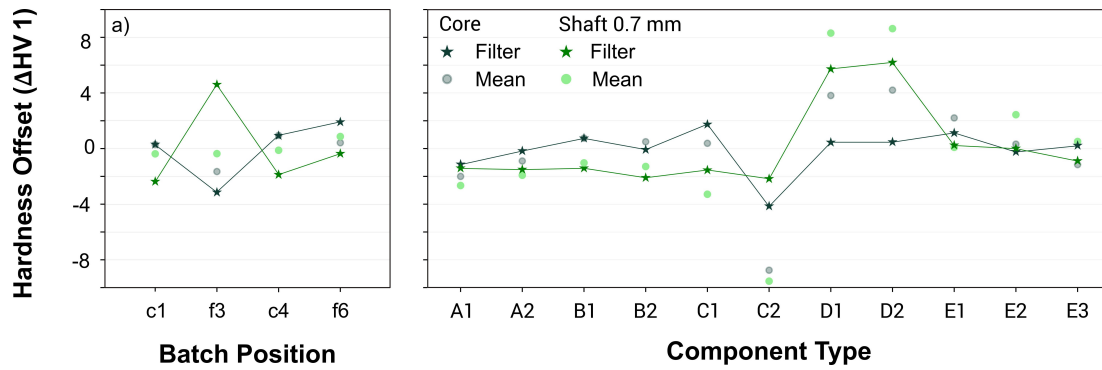
Figure 6. Offset in hardness at the core (dark green) and the Shaft 0.7 mm from the surface (light green) due to a) batch position and b) component type. Offsets were determined by i) filter model (star) and ii) by calculating the mean hardness (circle) for all nozzle bodies belonging to a given category.

Bild 6. Härteabweichung am Kern (dunkelgrün) und am Schaft 0,7 mm von der Oberfläche (hellgrün) aufgrund von a) Chargenposition und b) Bauteiltyp. Die Offsets wurden bestimmt durch i) das Filtermodell (Stern) und ii) die Berechnung der mittleren Härte (Kreis) für alle Düsenkörper, die zur selben Kategorie gehören.

nents sit at the same batch position over time, thereby skewing the mean positional influence. On the other hand, not all components are produced in the same quantity over time, which means that components produced in a time with higher carbon content appear harder, while others are seemingly less hard. The filter approach solves this problem by separating the two influences and additionally corrects for the fluctuation by tracking the base hardness independent from position and type. The optimal updating coefficient on the training set (i.e., data before June 2018) was found to be $a = 0.90$, while the optimal offsets differ for each nozzle body type, Figure 6. The remaining results are calculated on the test set (i.e., data after June 2018).

To estimate how much variance is explained by the different tracking model parts (Equation 8), the results from five scenarios are compared: using (1) the complete model (all), (2) setting position and component offsets to zero (only filter), (3) and (4) setting either one to zero (without position or component), and (5) always predicting the mean of the label distribution (dummy model), *Table 3*. About two-fifths of the variance (i.e., $R^2_{Core} = 0.4$ and $R^2_{Shaft} = 0.44$) can be predicted by the models, with the largest contribution coming from tracking current hardness (i.e., mainly fluctuations due to changes in material composition). Not knowing the component type does not lead to major losses in predictability (nozzle body diameter only varies slightly). Although the position is of some sig-

Table 3. Test set prediction results from tracking model in form of root mean squared error (RMSE) and coefficient of determination (R^2) for different scenarios (i.e., including and excluding batch position and/or component type in prediction).

Tabelle 3. Vorhersageergebnisse des Tracking-Modells für den Testdatensatz in Form des mittleren quadratischen Fehlers (RMSE) und des Bestimmtheitsmaßes (R^2) für verschiedene Szenarien (d.h. einschließlich und ausschließlich der Chargenposition und/oder des Komponententyps bei der Vorhersage).

		All	Only filter	Without position	Without component	Dummy model
RMSE	Core	9.28	9.48	9.48	9.29	12.13
	Shaft 0.7	8.04	8.88	8.51	8.44	10.72
R^2	Core	0.40	0.370	0.371	0.39	0
	Shaft 0.7	0.44	0.314	0.369	0.38	0

nificance (e. g., R^2_{Shaft}: all $=0.440$ vs. without position $=0.369$), its influence on the overall variance is still comparatively small. Thus, when comparing hardness measurements of two batches close in time, it may be more important to know the specimen position and type. Otherwise, material composition or overall drift may better explain discrepancies, assuming that the hardness measurement can be trusted, which is often enough only partially the case.

3.3 Process influence

Since industrialized heat treatment is a tightly controlled process, the measured parameters only vary by a very small amount and more strongly between furnaces than between consecutive batches in the same furnace, *Figure 7*. It is important to notice that these values stem from the sensor elements in the furnace that are also used to control the process. That means, if a sensor element itself has an offset to the true temperature (e. g., by calibration or remounting during maintenance) this deviation is not captured by the feature. Based on these minimal variance features (described in Section 2.3), none of the machine learning algorithms were able to make predictions on the test set with an R^2 score significantly different from zero, neither for the chronological nor the random data split. Overfitting of the training data was counteracted by reducing the model size (e. g., less nodes or trees in the arti-

ficial neural network and random forest). Training on individual furnaces also did not produce useful results, nor did correcting for long-term feature drifts using slow filters. In summary, this means that, either the process itself does barely contribute to the overall variance, or we did not or cannot measure the influencing factors.

3.4 Measurement error

Measuring hardness with a single HV 1 indentation is prone to considerable scatter, with about one in five measurements deviating further than 5 HV 1 from the estimated true mean, *Figure 8*. The RMSE of 4.2 HV 1 for the Core and 4.7 HV 1 for the Shaft 0.7 (not significantly different with $p=0.055$ by *F*-Test) provide a lower bound for the noise that is introduced to the overall distribution by the measurement procedure. It is a lower bound because the diamond replacement in 2018 resulted in measurements with less scatter, due to sharper edges of the indention. Consequently, the measurement optic can detect more precisely. For the measured hardness values, an increase in the diagonal length of 0.1 μm already leads to a decrease of 1.5 HV 1, being the most accurate that the optics can just about distinguish [19]. Measurements used for error estimation (i. e., four indents per measurement position) were made in 2020. Therefore, about two fifths of the original distribution contain
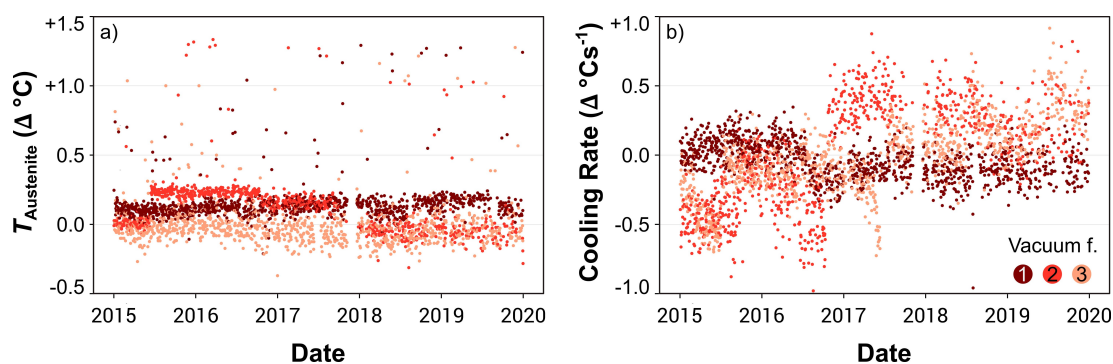

Figure 7. Features extracted from the three vacuum furnaces scaled around their respective reference value where a) is the mean austenitizing temperature, and b) is the mean cooling rate between 800 °C and 500 °C. Each dot represents the heat treatment of one batch.

Bild 7. Aus den drei Vakuumöfen extrahierte Merkmale, skaliert um ihren jeweiligen Referenzwert, wobei a) die mittlere Austenitisierungstemperatur und b) die mittlere Abkühlungsrate zwischen 800 °C und 500 °C ist. Jeder Punkt steht für die Wärmebehandlung einer Charge.
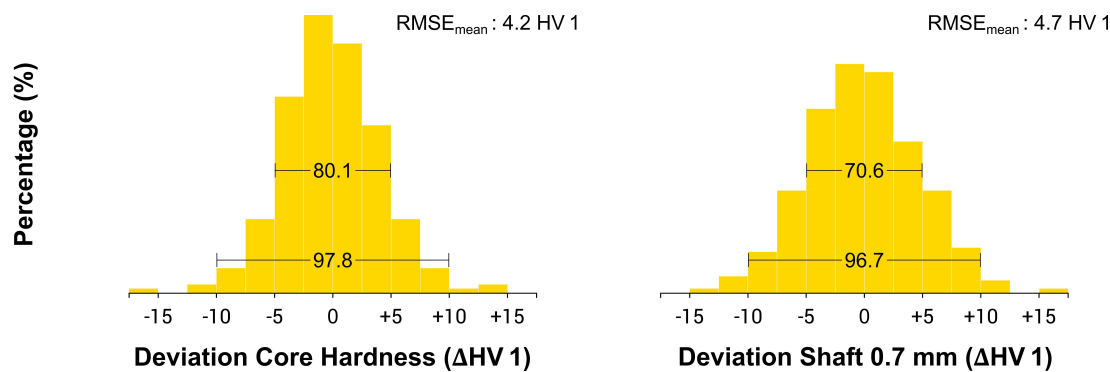
Figure 8. Distribution of the respective deviation from the mean value of four HV 1 indents on a single nozzle body per measuring position (i. e., core and stem 0.7 mm) for 68 test specimens.

Bild 8. Verteilung der jeweiligen Abweichung vom Mittelwert von vier HV 1-Eindrücken auf einem einzelnen Düsenkörper pro Messposition (d. h. Kern und Schaft 0,7 mm) für 68 Prüfkörper.
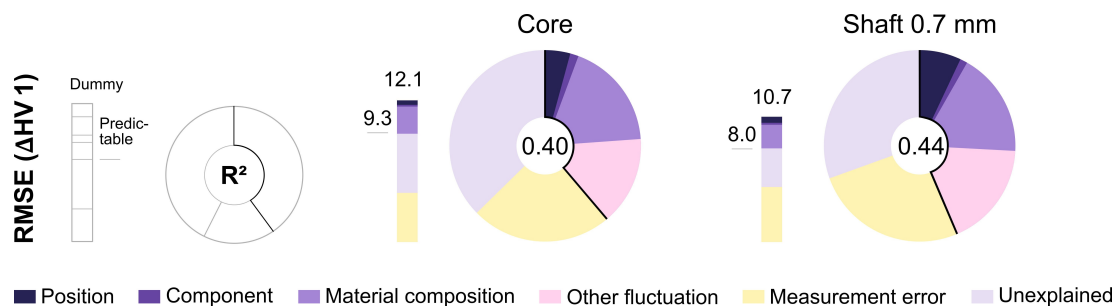

Figure 9. Contribution to RMSE and variance of several independent variables to hardness at the core and Shaft 0.7 mm.

Bild 9. Beitrag zum RMSE und zur Varianz verschiedener unabhängiger Variablen zur Härte im Kern und am Schaft 0,7 mm.

values with an even higher measurement error, leading to a larger RMSE contribution.

4 Summary

The overall variance of core hardness and shaft 0.7 is due to many factors. Some of which can be predicted, others only measured. An estimate of their relative contribution to the variance is based on their potential as a predictive variable (i. e., increase or decrease in R^2 when the variable is added or removed) and the amount they add to the RMSE, *Figure 9*. These estimates should not be taken at face value, but as a rough estimate and first start when aiming to reduce variance and estimate how much can be gained by this step. Influences of batch position and component type are almost irreducible but small (e. g., although small batch sizes may increase batch homogeneity, it would be eco-

nomically unreasonable). In stable high-volume industrial case hardening processes, material composition matters with miniscule changes in carbon leading to visible variations in the resulting core hardness. While the mid-term fluctuations could be readily explained by material composition (i. e., Maynier's model), a tracking approach employing filters is better suited for prediction of the current core hardness level, as the model can also follow drifts from the measurement device as well as make position and component type specific predictions. Surface near measuring positions are much less affected by composition of the base material due to carburization by acetylene during heat treatment. Probably the largest variance contributor is the HV 1 measurement itself, due to the small indent size. Using HV 10 measurements is not feasible at the surface near regions because a defined distance of the indent to the boundary must not be fallen short of and changing test load between measurements

likely introduces operator mistakes. Multiple indentations are a feasible measure but may be economically undesirable.

Acknowledgements

5 References

[1] M.J. Faizabadi, G. Khalaj, H. Pouraliakbar, M.R. Jandaghi, *Neural Comput & Applic* **2014**, *25*, 1993; Applic **2014**, 25, 1993.

[2] P.H. Maynier, J. Dollet, P. Bastien, *Metall. Soc. AIME* **1978**, 163.

[3] P. Maynier, B. Jungmann, J. Dollet, *Hardenability Concepts Appl. Steel, Proc. Symp.* **1977**, 518.

[4] G. Sidhu, S.D. Bhole, D.L. Chen, E. Essadiqi, *Mater. Des.* **2012**, *41*, 99.

[5] W.G. Vermeulen, P.J. van der Wolk, A.P. de Weijer, S. van der Zwaag, *J. Mater. Eng. Perform.* **1996**, *5*, 57.

[6] H. PourAsiabi, H. PourAsiabi, Z. AmirZadeh, M. BabaZadeh, *Mater. Des.* **2012**, *35*, 782.

[7] J. Trzaska, L.A. Dobrzański, *J. Mater. Process. Technol.* **2005**, *164–165*, 1637

[8] Z. Sterjovski, D. Nolan, K.R. Carpenter, D.P. Dunne, J. Norrish, *J. Mater. Process. Technol.* **2005**,*170(3)*, 536

[9] M. Rahaman, W. Mu, J. Odqvist, P. Hedström, *Metall. Mater. Trans. A* **2019**, *50*, 2081.

[10] G. Sidhu, S.D. Bhole, D.L. Chen, E. Essadiqi, *Comput. Mater. Sci.* **2011**, *50*, 3377.

[11] C. Garcia-Mateo, T. Sourmail, F.G. Caballero, C. Capdevila, C. García de Andrés, *Mater. Sci. Technol.* **2005**, *21(8)*, 934.

[12] C. Bailer-Jones, H. Bhadeshia, D. MacKay, *Mater. Sci. Technol.* **1999**, *15(3)*, 287.

[13] J. Pak, J. Jang, H. Bhadeshia, L. Karlsson, *Mater. Manuf. Processes* **2008**, *24 (1)*, 16.

[14] S.B Singh, *Ironmaking Steelmaking* **1998**, *25(5)*, 355.

[15] H. Jin, S. Wu, Y. Peng, *J. Mater. Eng. Perform.* **2013**, *22(12)*, 3631.

[16] Y. Lingelbach, L. Hagymási, T. Waldenmaier, V. Schulze, *HTM, J. Heat Treat. Mater.* **2020**, *75*, 212.

[17] A. Diemar, *Ph.D. Thesis*, Bauhaus-Universität Weimar, Germany, **2007**.

[18] A. Horsch, presented at *Härtereikreis*, Mittlerer Neckar, Germany, March 07, **2017**, pp. 1–6.

[19] Norm, Beuth, Berlin, DIN EN ISO 6507–3, **2018**.

[20] J. Trzaska, *Arch. Comput. Mater. Sci. Surf. Eng.* **2013**, *61*, 87.

[21] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, İ. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, *Nat. Methods* **2020**, *17*, 261.

[22] M.V. Li, D.V. Niebuhr, L.L. Meekisho, D.G. Atteridge, *Metall Mater Trans B* **1998**, *29*, 661

[23] N.S. Reddy, J. Krishnaiah, S.-G. Hong, J.S. Lee, *Mater. Sci. Eng. A* **2009**, *508*, 93.

[24] J. Tenner, *Ph.D. Thesis*, University of Sheffield, England, **2000**.

[25] M. McCartney, M. Haeringer, W. Polifke, *J. Eng. Gas Turbines Power* **2020**, *142*.

[26] E. Bélisle, Z. Huang, S. Le Digabel, A.E. Gheribi, *Comput. Mater. Sci.* **2015**, *98*, 170.

[27] M. Kuhn, K. Johnson, *Applied predictive modeling*, Springer, New York **2013**.

[28] C.M.Bishop, *Pattern recognition and machine learning*, Springer, New York **2006**.

[29] T. Head, M. Kumar, H. Nahrstaedt, G. Louppe,I. Shcherbatyi, Zenodo, **2020**.

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P.

Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*,

[31] J. Reunanen, *J. Mach. Learn. Res.* **2003**, *3*, 1371.