Hybrid Modelling of Austempering in the Automotive Industry

Jonathan Wörner[1], Thomas Waldenmaier[2], László Hagymási[2] and Volker Schulze[3]

[1]Robert Bosch GmbH, Stuttgart-Feuerbach 70469, Germany
[2]Robert Bosch GmbH, Renningen 71272, Germany
[3]Institut für Produktionstechnik (wbk), Karlsruher Institut für Technologie (KIT), Karlsruhe 76131, Germany

Austempering is a heat treatment process used in the automotive industry. In contrast to case hardening, austempering results in a bainitic microstructure with only slightly lower hardness and significantly higher fatigue strength. For economic reasons, the heat treatment is carried out in batch processes, usually with several components being quality tested.

Extensive sensor data is available for each batch, but the variance is low due to the stability of the series process. At the same time, quality control using Vickers hardness testing is subject to measurement uncertainty, especially when only a few indentations are measured, as is common in production. Deviations within a batch further limit the prediction accuracy for an individual part. The prediction quality of a machine learning (ML) model is limited by the quality of the data. One way to better model the process, is to use simulation data in a hybrid model. Another approach is Uncertainty Quantification, which provides an indication of how confident the model is that the prediction is accurate. Many of these models are also more robust and have higher prediction accuracy.

The aim of this work is the implementation of several Uncertainty Quantification methods for a hybrid model as well as the comparison of these models and a recommended action for hybrid models in production. The ML models with Uncertainty Quantification are better suited for data with a higher uncertainty (e.g. measurement uncertainty). These models learn next to the labels how reliable the predictions are. This is particularly important in production and can be used as an additional criterion in quality control. The developed model will be evaluated with production data to assess its performance.

Keywords: *austempering, hybrid modeling, artificial intelligence, bayesian optimization*

1. Introduction

The digitization of manufacturing in the automotive industry is generating large amounts of data that can be used to optimize and reduce costs in production. In particular, the use of machine learning (ML) and deep learning algorithms offers great potential for increasing efficiency and quality in production.

Constant processes and the associated low variance of the process data are a major challenge for the use of ML in production. This is especially the case in the heat treatment of metallic materials, where processes are developed for a product and then little or no adjustments are made over the years. An example of this is the austempering process. Internal investigation have shown, that unlike martensitic hardening, austempering produces a bainitic microstructure with only slightly less hardness but significantly better fatigue strength.

For economic reasons, heat treatment is carried out in batch processes, where usually only a few individual components of a batch are subjected to quality testing. This involves metallographic testing. This inspection is costly because it is destructive and requires expert knowledge.

An important test criterion for the quality of the batch is the hardness measurement of test specimens. At the same time, quality control with hardness testing is subject to measurement uncertainties, especially when only a few indentations are measured, as is often the case in production. One method of hardness testing is the Vickers hardness measurement, which is used in this work. The prediction quality of a ML model is limited by the quality of the data. As hardness measurement shows scatter in the results, it is crucial to check the uncertainty of the prediction. By uncertainty quantification, it is possible to assess whether the prediction is valid. Standard ML models cannot represent these uncertainties, nor do they understand the limits within which they are valid.

Therefore, this paper presents two methods for uncertainty quantification, implements a hybrid model for Vickers hardness prediction, and shows the advantages and disadvantages of the applied methods. The prediction model is a ML model that is additionally fed with physical information. Through the combination of simulation and ML techniques in a hybrid model, model properties such as robustness and explainability can be improved (1).

The focus of this work is to explore uncertainty quantification in the context of predicting Vickers hardness measurements for an austempering series process. The rest of the paper is organized as follows. In Section 2, a literature review of the uncertainty quantification and the background of the austempering process is given. In Section 3 the Results of the uncertainty quantification in the context of the austempering is discussed and in Section 5 a conclusion is provided.

2. Methodology and Process

2.1 Uncertainty Quantification (UQ):

The application of ML methods in manufacturing and materials science is becoming increasingly popular. Most recently published ML approaches are deterministic, i.e., for each input there is a unique output (2). However, deterministic models have limited understanding of their own knowledge (3) and often overestimate (4), which can lead to high error costs in production. Therefore, accounting for uncertainty in ML models can increase the reliability of prediction results (5).

In uncertainty quantification, two different uncertainties can be distinguished, Aleatoric and Epistemic Uncertainty. *Aleatoric Uncertainty*, also called data uncertainty, describes the uncertainty which results from noise in the data (e.g., measurement uncertainty). This uncertainty will not be further investigated in this work.

Epistemic Uncertainty (EU), on the other hand, describes uncertainty due to missing information in the data and can be reduced by adding the missing data (6, 7). In addition, the EU can identify situations that were not part of the training dataset (5). The EU is calculated by measuring the variance in the predicted means (8). To illustrate this uncertainty, consider an example from the field of autonomous driving. Suppose a system is trained in Tokyo (Japan) and then the car is driven to Stuttgart (Germany). In Stuttgart, we drive on the right side of the road and the signs look different than in Tokyo. Here the system should realize that its knowledge is limited and cannot be used for this situation (6).

Uncertainty quantification for regression can be divided into the *Statistical* and the *"Bayesian"* approaches. The Statistical approaches contain the methods S*ingle Hypothesis*, L*ikelihood Estimation* and *Density Estimation*. The "Bayesian" approaches includes methods like *Gaussian Processes* and *Bayesian Neural Networks (BNN*) (3). In this work the focus will be on the likelihood estimation, more specific the *Fisher Information (FI)* and the BNN.

The *Maximum Likelihood Estimation (MLE)* is a statistical principle that involves finding the parameter values of a model that maximize the probability of observing the given data and serves as a common method for parameter estimation. FI quantifies the curvature of the likelihood function. This provides information about the precision of parameter estimates and the confidence intervals, which is related to EU (3). Using the described method the confident value is calculated using the ForestCI Package in python (9).

BNN on the other hand are neural networks that model the EU by representing weights as probability distributions, also called prior. Instead of optimizing for fixed weights, Bayesian inference is used to compute the posterior distribution over the weights. This posterior represents a range of plausible model parameters given the data (3). Additional the Aleatoric Uncertainty is modeled as a Gaussian distribution with added observation noise in the last neural network layer (5).

For the prediction of the uncertainty the ML models are often over or underconfident, for this reason the uncertainty has to be calibrated prior to the prediction (10). For the calibration of the uncertainty the python package Uncertainty-Toolbox was used (11).

2.2 Process Chain & Quality Assurance

The process under consideration is a two-stage austempering process for automotive components. The components are heat treated in batches in a multi-chamber furnace. A batch consists of several hundred components made of the material 100Cr6 (ASTM 52100), which are arranged in several layers.

The components are manufactured in series on ten matching IPSEN TQA-4 (5) chamber furnaces. These furnaces are referred to as salt bath lines. The multi-chamber furnaces consist of three stations, a process gas furnace, a salt bath, and a low temperature convection furnace. In the first furnace, the components are fully austenitised with partial dissolution of carbides. In the salt bath, the components are quenched and held just above the martensite start temperature. In the last furnace, further transformation to bainite takes place at a temperature above the salt bath. The components are held in the convection furnace until the desired degree of bainite transformation is achieved. Finally, the parts are cleaned and inspected in the metalographic lab to ensure quality.

For quality control, one or two components are taken from each batch at specified locations. These are tested as representative of the batch. In addition to further metallographic examinations, hardness is measured at a specified position on the surface and in the core of the component. The hardness measurement is Vickers using HV10. Three hardness indentations are made for each position in accordance with the standard and the average value at each position is stored in a database.

For traceability and process monitoring and control, data is collected for each batch and also stored in a database. All data can be assigned to a batch using a unique ID. Metadata such as furnace number and process start are stored in addition. For process monitoring, sensor signals of the temperature in the furnace and the gas flow of the inert glass are also recorded.

Before extracting features from described data, the data must be prepared. This means removing incorrect and missing data as well as unifying timestamps. To extract the features, the metadata is transformed into columns with values of one and zero using *One-Hot Encoding*. For each furnace chamber, the duration is determined from the transient data and characteristic segments are identified. For each segment, the mean, minimum, maximum and skewness are calculated. For the hybrid approach the integral over the Hollomon-Jaffe parameter was calculated for each segment. In addition, the features were extended by the results of Erick Dabrock's empirical equation for dry austempering (12).

In addition, feature selection for the BNN is performed as a function of the *F-score* to reduce the number of features.

The core hardness is used as a label for the models. The hardness is subject to uncertainties due to the measurement method. For example, according to DIN-EN-ISO-6507-3, with a reference hardness of 700 HV, a maximum range of 3.4 %, or 23.8 HV, is allowed for 25 HV10 measurements on a refence plate. Assuming that the measurements are normally distributed, this error corresponds to a standard deviation of about 4.1 HV (13). For serial measurements, the error can be expected to be higher.

The MLE is applied to *Forest Ensemble Regressors (FER)*. With the best result for the *Random Forest Regressor (RFR)* and the *Extra Trees Regressor (ETR)*. To find the optimal parameters of the regression model, a *Bayesian Hyperparameter Optimization (BHPO)* was performed on the R^2 metric.

3. Results and Discussion

Series data was used to train the models. This data contains the process data of different types of components. These components were heat treated with variations in the process. The data is not divided randomly, but all the data before a certain point in time form the training data and all the data after the point in time form the test data. If the data were split randomly, it could happen that the model performed well on the test data set but poorly in predicting future batches. The data sets are divided by days, where one day corresponds to an average of two digits of produced parts. 30 days were chosen for the test data and 10 days for the validation data.

Table 1: Result Hyper-Parameter-Optimization for ML algorithm with UQ.

Algorithm	Toolbox	Hyperparameter
ETR	Scikit-learn	n estimators: 2200, max depth: 20, bootstrap: True, train data: 547 days
BNN	Tensorflow	input layer (231), hidden layer (10) output layer (2), train data: 284 days

Hyperparameter optimization was performed to identify the best model for the FER and BNN models. Optimization for the FER models are much simpler due to the smaller number of parameters and training time. While training a FER model takes a few minutes, training a BNN model takes several hours on a GPU cluster. The best model and parameters for the FER and BNN models are shown in Table 1. Table 2 shows the accuracy metrics of the two best models. The accuracy of the models is compared using the coefficient of determination R2, which indicates the proportion of scatter in the underlying distribution, and the mean absolute error (MAE). Uncertainty is reflected by the Mean Absolute Calibrated Error (MACE) metric. The ETR model performs best for both accuracy metrics. Considering the expected measurement uncertainty of the hardness test from the previous chapter and comparing it with the MAE, it is noticeable that the ETR model with an MAE of 4.41 HV is close to the value of 4.1 HV from the standard. The resulting deviation can be explained by the serial process, as it is not in the optimum conditions for a reference plate.

Table 2: Core hardness prediction results with uncertainty.

	ETR + MLE	BNN
R^2	0.89	0.86
MAE	4.41	4.97
MACE	0.02	0.04
EU (mean)	5.71	5.35

Figure 1 shows how well the two models represent the hardness of the components. In this figure, the abscissa shows a section of the validation data labels (core hardness) and prediction of the models sorted by value. The ordinate shows the relative distribution of the hardness values. While the BNN model shows different levels that can be matched to different products, the ETR model provides a much better representation of hardness across the label space. Furthermore, the calibrated EU (in bright blue) describes the uncertainties of the ETR model well, with the uncertainty increasing as the predicted value deviates from the measured value. This correlation is also shown by the MACE metric.

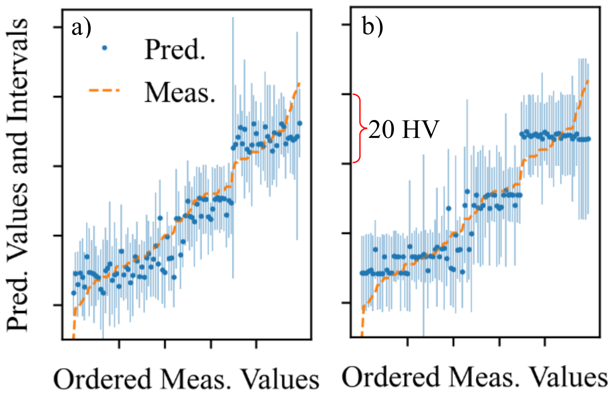

Figure 1: Ordered Prediction Intervals: a) ETR-Model and b) BNN-Model.

4. Conclusion and Outlook

Large amounts of process and test data are stored in series production. To use the data to predict test criteria such as hardness, data must be processed, and relevant features extracted. Low variation of the data combined with high measurement uncertainty is a challenge. ML models with UQ are particularly suitable for this type of data. Not only do they represent the data well, but they also provide an estimate of the quality of the prediction. This is particularly useful in manufacturing environment. When deciding whether to inspect a batch, not only the prediction result can be used, but also how confident the model is in its prediction. While the ETR model outperformed the BNN model on the data shown, the BNN model offers advantages when considering the aleatory uncertainty. Which is not possible with the FER models. In the future, models with UQ could be used to predict test results to reduce the frequency of testing.

3. Reference

5. [1] KURZ, Stefan, Herbert de GERSEM, Armin GALETZKA, Andreas KLAEDTKE, Melvin LIEBSCH, Dimitrios LOUKREZIS, Stephan RUSSENSCHUCK, and Manuel SCHMIDT. Hybrid modeling: towards the next level of scientific computing in engineering [online]. *Journal of Mathematics in Industry.* 2022, **12**(1). Available from: 10.1186/s13362-022-00123-0.

[2] KANEKO, Tatsuya, Ryota WADA, Masahiko OZAKI, and Tomoya INOUE. Hybrid physics-based and machine learning model with interpretability and uncertainty for real-time estimation of unmeasurable parts [online]. *Ocean Engineering.* 2023, **284**, 115267. Available from: 10.1016/j.oceaneng.2023.115267.

[3] HÜLLERMEIER, Eyke, and Willem WAEGEMAN. Aleatoric and Epistemic Uncertainty in Machine Learning: An Introduction to Concepts and Methods [online]. 2019. Available from: 10.48550/arXiv.1910.09457.

[4] ABDAR, Moloud, Farhad POURPANAH, Sadiq HUSSAIN, Dana REZAZADEGAN, Li LIU, Mohammad GHAVAMZADEH, Paul FIEGUTH, Xiaochun CAO, Abbas KHOSRAVI, U. Rajendra ACHARYA, Vladimir MAKARENKOV, and Saeid NAHAVANDI. A review of uncertainty quantification in deep learning: Techniques, applications and challenges [online]. *Information Fusion.* 2021, **76**, 243-297. Available from: 10.1016/j.inffus.2021.05.008.

[5] KENDALL, Alex and Yarin GAL. *What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?:* arXiv, 2017.

[6] MALININ, Andrey. *Uncertainty Estimation in Deep Learning with application to Spoken Language Assessment,* 2019.

[7] PHAN, Buu Truong. *Bayesian Deep Learning and Uncertainty in Computer Vision:* UWSpace, 2019.

[8] DEPEWEG, Stefan, José Miguel HERNÁNDEZ-LOBATO, Finale DOSHI-VELEZ, and Steffen UDLUFT. *Decomposition of Uncertainty in Bayesian Deep Learning for Efficient and Risk-sensitive Learning:* arXiv, 2017.

[9] POLIMIS, Kivan, Ariel ROKEM, and Bryna HAZELTON. Confidence Intervals for Random Forests in Python. *Journal of Open Source Software.* 2017, **2**(1).

[10] ZHAO, Shengjia, Tengyu MA, and Stefano ERMON. *Individual Calibration with Randomized Forecasting:* arXiv, 2020.

[11] CHUNG, Youngseog, Ian CHAR, Han GUO, Jeff SCHNEIDER, and Willie NEISWANGER. *Uncertainty Toolbox: an Open-Source Library for Assessing, Visualizing, and Improving Uncertainty Quantification,* September 2021.

[12] DABROCK, Eric. *Experimentelle und modellbasierte Untersuchung der Wirkzusammenhänge zwischen Prozessparametern und mikrostrukturellen sowie mechanischen Eigenschaften am Beispiel des trockenen Bainitisierens von 100Cr6.* 1. Auflage. Kaiserslautern: Universität Kaiserslautern, 2020. Schriftenreihe der Arbeitsgruppe Werkstoffprüfung. 5. 978-3-95974-126-2.

[13] *DIN EN ISO 6507-3:2018-07, Metallische Werkstoffe_-Härteprüfung nach Vickers_- Teil_3: Kalibrierung von Härtevergleichsplatten (ISO_6507-3:2018); Deutsche Fassung EN_ISO_6507-3:2018.* Berlin: Beuth Verlag GmbH.