Modeling of surface hardening and roughness induced by turning AISI 4140 QT under different machining conditions

Benedict Stampfera,*, Jeanine Bachmannb, Daniel Gaudera, David Böttgerb, Michael Gerstenmeyerb, Gisela Lanzaa, Bernd Wolterb, Volker Schulzea

Abstract

For this work, the surface-layer states of turned AISI 4140 QT were investigated by means of surface roughness and microhardness measurements. Different machining conditions are regarded, namely cutting velocity, feed rate, tool wear and the tool corner radius, as well as the tempering state of the workpiece. The resulting data is analyzed by multiple algorithms, in order to create analytical models for a real time process control. Modeling approaches applied are linear regression, stepwise regression, LASSO and Elastic Net. Finally, the models are evaluated in terms of quality, complexity and physical plausibility.

Keywords: Process modeling; Hardness; Roughness

1. Introduction

White layers are microstructural changes, which can result from machining of quenched and tempered steels. Those layers can be thermally [1, 2] or mechanically [3] induced. Softer dark tempering zones may also occur. Depending on the type of modification and the application, those layers can be detrimental to the function of the machined component. In order to avoid this and control to surface states during machining, an understanding of the underlying mechanisms is essential. In a previous work, hardness distributions and the associated microstructural changes were recorded and evaluated by micrographs and microhardness tests for AISI 4140 quenched and tempered (QT) [4]. The results were analyzed with respect to prevailing process forces. It was concluded that passive force was the most promising indicator of mechanical stress and the resulting heat input into the workpiece. Likewise, the feed rate was identified as an important source of influence, as well as that reducing the feed rate lowers the passive force and thus reduces surface hardening.

Quantitative evaluation methods and models are particularly valuable for the realization of in-process controls for surface integrity. In general, surface integrity is defined by a variety of parameters, including chemical composition, microstructure, dislocation density, hardness, local strength, residual stress tensor, and topography [5] which result from process characteristics and may improve the quality of machined components. In the following section, concepts and models for the control of process characteristics and surface-layer states are reviewed.

Gauder et al. [6] investigated turning of AISI 4140 QT and used multivariate regression to determine the relationships between the process forces as a function of feed rate, cutting speed, tool corner radius, cooling strategy and depth of cut on specimens with tempering temperatures of 300, 450 and 600°C. They determined all variables to be significant and reached a
good prediction accuracy based on the coefficient of determination $R^2$ and the root mean squared error RMSE.

Böttger et al. [7] describe a soft-sensor concept for external longitudinal turning of AISI 4140 QT for the prediction of relevant surface-layer modifications such as white layers during machining. A wear mark model, based on process forces and acoustic emission, as well as a component model, based on measured micromagnetic material parameters testing technique plays a major role. These measurements correlate with the material microstructure, can be performed in the machining area and could provide data for an in-process control in further investigations.

For orthogonal cutting of AISI 4140 QT, Meurer et al. [8] investigated the dynamic recrystallization and the formation of white layers, as well as the prediction of their occurrence. For this purpose, an analytical model of the process forces was used to estimate the resulting temperature fields and to evaluate the correlation with a validated chip formation simulation. This was then developed into a soft-sensor model to predict white-layer formation during the process. The resulting model was validated using micrographs.

Sada [9] investigated the application of a neural network to model the surface condition of mild steel after turning. As part of this, he used the process parameters cutting speed, depth of cut and the feed for his model to predict the material removal rate and the average surface roughness $R_a$. To create the neural network, he used a training dataset of 40 observations and the Levenberg-Marquart and Sealed Conjugate Gradient algorithms. $R^2$ and RMSE were consulted as evaluation criteria and the results were compared with previous regression analyses. It was found that the Levenberg-Marquardt algorithm with an optimum of ten hidden neurons achieved an accurate and effective prediction of the experimentally collected data.

Zemzemi et al. [10] investigated the prediction of occurring white layers during orthogonal turning of AISI 52100 hardened steel by studying the cutting forces and the thermomechanical effects on the machined surface and the mechanical stresses. To predict the cutting forces, they combined the Oxley and Waldorf models, the theory of obliquely moving strip heat sources for the temperature distribution, and the rolling/sliding approach for the mechanical stresses. They concluded that their analytical results agree well with the literature. The thickness of the white layers increases with cutting speed, feed rate and tool wear while the thickness of the tempering zone shows a decreasing trend.

Kuntoğlu et al. [11] investigated the influence of cutting speed, feed rate and cutting edge angle on vibration during turning of AISI 5140 and roughness after turning. Response surface methodology, ANOVA based analysis and quadratic regression models were used to optimize the vibration and surface roughness. This gave best results for the parameters cutting speed $v_c = 190$ m/min, feed rate $f = 0.06$ mm and rake angle $\kappa = 60^\circ$. Using regression, a model was created that predicted the roughness and vibrations with a high accuracy and was validated by means of further experiments. These showed errors of less than 10%.

Uhlmann et al. [12] investigated the influence of tool wear and machining parameters when milling AISI 4140 QT. Due to the strong correlation of the flank wear and the averaged motor module power, it was concluded that power measurements show a great potential for obtaining the tool wear in-process. Besides this, the maximum Barkhausen noise amplitude and the coercive force of machined specimens were measured. Characteristic limits for the prediction of surface damage were exceeded in tests with increasing tool wear. It was further shown that the magnetic parameters can be reduced by an adoption of machining parameters. Finally the magnetic properties were not measured but predicted by neural networks, random forest and gradient boost machine learning, based on machining parameters and the tool wear. The thereby reached coefficients of determination exceeded above 0.9.

Glatt et al. [13] predicted the martensite content of AISI 347 after cryogenic turning by machine learning models, using the process forces and the workpiece temperature as inputs. From the tested algorithms random forest, neural network and support vector machine, the latter led to the most accurate results in terms of RMSE and $R^2$ while needing the lowest computational resources.

It can be concluded that the model based control of surface states in cutting is established more and more in scientific practice. The aim of the present article is the quantitative description of surface states after longitudinal turning of the quenched and tempered steel AISI 4140 QT. Machine learning algorithms such as support vector machines and neural networks have been applied successfully for the prediction of surface states. However for those models, the estimation of the physical plausibility and the determination of quantitative model sensitivities to specific input parameters is challenging, which impedes the generation of mechanism based process knowledge. Linear regression of polynomial approaches is a widely used alternative, which generally allows to estimate the physical plausibility and parameter sensitivities. Yet, the problem of overfitting is rarely addressed in literature with these approaches. Hence, this will be focused in the present work.

2. Experimental Setup

The investigated cylindrical shafts had a length of 200 mm and an initial diameter of 48 mm. Before cutting, the AISI 4140 specimens were quenched according to DIN EN ISO 683-2 and tempered for 1 h at 300, 450, 600 or 640 °C. As presented in [4], the longitudinal turning tests were conducted on an index G200 machine type. The experimental setup includes a tool holder of the type SCLCR202K12 and TiCN-coated carbide inserts of the type Walter CCMT120404/12 RP4 WPP20S. The parameters that were kept constant during the experiments are the clearance angle $\alpha = 7^\circ$, the macroscopic rake angle $\gamma = 0^\circ$, the principal cutting edge angle $\kappa = 95^\circ$ and the cutting edge inclination $\lambda = 0^\circ$. The varied machining parameters and the respective levels are presented in Table 1. The tests were conducted in multiple runs and did not follow a rigid design of experiments. This contributes to the needs of process modeling.
in non-scientific environments with quasi statistical datasets, which still can be restricted to ranges of interest. Instead of explicitly defining tool wear levels, the wear evolution was measured between the cuts by an optical microscope, resulting in the specified range. The tool edge radius was included into the flank wear mark VB, leading to an initial value VB = 0.075 mm. An emulsion was used for process cooling, consisting of water and 10% oil of the type Motorex Swisscool 8000.

Table 1. Varied parameters and insert geometries for the turning tests.

<table>
<thead>
<tr>
<th>parameters</th>
<th>levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>tempering temperature QT</td>
<td>300, 450, 600, 640 °C</td>
</tr>
<tr>
<td>tool corner radius r_c</td>
<td>0.4; 1.2 mm</td>
</tr>
<tr>
<td>Tool wear VB</td>
<td>0.075 – 0.4 mm</td>
</tr>
<tr>
<td>cutting velocity v_c</td>
<td>100, 200, 250, 300 m/min</td>
</tr>
<tr>
<td>feed f</td>
<td>0.05; 0.1; 0.15; 0.2; 0.3 mm</td>
</tr>
<tr>
<td>depth of cut a_p</td>
<td>0.2; 0.3; 0.4; 0.6 mm</td>
</tr>
<tr>
<td>Cooling strategy</td>
<td>dry, emulsion 10%</td>
</tr>
</tbody>
</table>

The machined surface topography was measured by a confocal light microscope of the type Nanofocus μsurf custom. The average surface roughness Ra, and the average maximum profile height Rz were then evaluated in longitudinal specimen direction with a cutoff length of 0.8 mm according to DIN EN ISO 4287. The roughness was measured for 77 specimens. The parameter distribution will be addressed in section 4.1.

The hardness HV0.005 was measured in transversal cross sections of the turned shafts, see also [4]. When turning AISI 4140 QT tempered at 300 °C, possible annealing effects reduce the hardness, while phase transformations can lead to a strong hardness increase. Turning AISI 4140 QT tempered at 450 °C or above leads to grain refinement, work hardening and thus a gradual hardness increase [4, 14]. In order to avoid severe nonlinearities, the specimens tempered at 300 °C were excluded from the hardness model database. For the same reason, only dry cutting experiments were considered. The Vickers indentations placed in surface distances from 5 µm to 200 µm showed an in-depth material modification of less than 50 µm. This coincidences with complementary micrographs and modifications identified after drilling of AISI 4140 QT [15, 16]. Consequently, the mean difference AHV between 5 and 50 µm was taken as hardness model target value. For each surface distance, six repetitions were conducted. The hardness increase was evaluated for 47 specimens. The parameter distribution will be addressed in section 4.2.

3. Modeling

The goal of this work is the identification of robust quantitative models, which contain the process parameters as explicit inputs. This permits physical interpretations and the evaluation of target value sensitivities to parameter variations. The approaches used are multilinear and fully quadratic polynomials of all process parameters and their combinations. While the quadratic approach is capable of modeling nonlinear dependencies, the high number of degrees of freedom (DOF) bears the risk of overfitting. Additionally, the incorporation of multiplied process parameters leads to collinearity and thus arbitrary fitted constants. Due to those issues, the variable selection techniques Stepwise Regression (SWR), Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net were applied. The models were fitted by a 10-fold cross validation (cv.) technique. Consequently, the model performance measures RMSE and R² presented in section 4 are mean values resulting from out-of-sample estimations during cross validation. Since the physical units of the fitted polynomial constants are consistent with the units of the process parameters in Table 1, they are not given in section 4.

The cooling strategy was incorporated as 0 for dry cutting and 1 for emulsion cooling in the model equations.

The data analysis and the process modeling was realized using the programming language R [17] and the packages tidverse [18], caret [19] and glmnet [20]. The latter reference also provides detailed information of the algorithms used for parameter selection.

4. Results and Discussion

4.1. Modeling of surface roughness

Before modeling the surface roughness, the distribution and the correlation of the dataset was analyzed by the help of a pairwise plot matrix, which is depicted in Figure 1. In the lower section of the matrix, the parameters are pairwise compared and the data dots are placed in the space spanned by the respective parameters. To improve the size and the readability of the matrix, the vertical scales are omitted. However, the range and the numbers are identical to the labels on the bottom for the respective parameters. The diagrams on the main diagonal depict the parameter levels and the distribution of the 77 observations.

![Fig. 1. Pairwise correlation matrix for the roughness dataset](image-url)
The regression lines in the diagrams of the lower matrix section and the factors in the upper section indicate parameter correlations. An almost ideal linear correlation is present between the average roughness \( R_z \) and the average maximum profile height \( R_s \). Consequently, one parameter fully explains the evolution of the other and a discussion of both is not necessary. A strong correlation between the process parameters would indicate an undesired imbalance in the dataset. As a counteraction, additional tests with underrepresented parameter combinations could be conducted. The correlations between the target values and the process parameters indicate a high impact of the feed on \( R_z \) and \( R_s \), which is be explained by equation (1) for the kinematic roughness [21].

\[
R_{z,\text{kinematic}} = r_x \left( 1 - \sqrt{1 - \left( \frac{r_x}{2r_z} \right)^2} \right) \approx \frac{r_x}{8r_z}
\]  

The negative correlation of \( r_x \), and \( VB \) with the roughness can be attributed to kinematic dependencies of tool and workpiece as well. The wear flattens the tool at the workpiece contact which lowers the roughness, similar to a high corner radius. The relatively strong correlation between the tempering state and the roughness surprises and could be an artefact of the correlations between \( QT \) and the process parameters \( r_x \) and \( VB \). Further it must be noted that correlation coefficients indicate a linear dependency, but not the respective slope. A physical explanation approach for a rougher surface is the side flow of higher tempered and thus more ductile workpiece material which was not present in the training set [22, 23]. LASSO and Elastic Net both generate better models, while Elastic Net algorithm tends to be more conservative. Both inputs increase mechanical surface loads and are known to fortify grain refinement. This could explain the strong correlation of \( QT \) and \( f \) with the roughness.

Table 2. Performance indicators for \( R_z \) models

<table>
<thead>
<tr>
<th>Model approach</th>
<th>Parameter selection</th>
<th>DOF after selection</th>
<th>10 fold cv. RMSE</th>
<th>10 fold cv. ( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>none</td>
<td>8</td>
<td>2.04 ( \mu m )</td>
<td>0.77</td>
</tr>
<tr>
<td>Quadratic</td>
<td>none</td>
<td>34</td>
<td>3.94 ( \mu m )</td>
<td>0.51</td>
</tr>
<tr>
<td>Quadratic</td>
<td>SWR</td>
<td>13</td>
<td>2.49 ( \mu m )</td>
<td>0.69</td>
</tr>
<tr>
<td>Quadratic</td>
<td>LASSO</td>
<td>15</td>
<td>1.76 ( \mu m )</td>
<td>0.82</td>
</tr>
<tr>
<td>Quadratic</td>
<td>Elastic Net</td>
<td>18</td>
<td>1.71 ( \mu m )</td>
<td>0.82</td>
</tr>
<tr>
<td>Data filter and LASSO</td>
<td></td>
<td>6</td>
<td>1.86 ( \mu m )</td>
<td>0.86</td>
</tr>
</tbody>
</table>

Without parameter selection, the quadratic approach performs worse than the linear. This shows the problem of classic polynomial regression, which is prone to overfitting, especially with a high number of DOFs. The quality of the model generated by SWR is as well worse than the linear approach, despite the relatively low number of selected DOFs. This confirms the shortcomings of SWR, when predicting data which was not present in the training set [22, 23]. LASSO and Elastic Net both generate better models, while LASSO requires fewer DOFs. LASSO is preconditioned to terminate parameters with a high collinearity, while Elastic Net algorithm tends to consider more parameters in order to reduce prediction errors. The LASSO model with 6 DOFs results from a database filter approach. It incorporates 50 instead of 77 observations and will be treated by the end of section 4.1. Disregarding the data filter approach, the LASSO model with 15 DOFs represents a good compromise, of few DOFs and a high prediction quality. The parameters determined are listed in Table 3.

Table 3. \( R_z \) model coefficients generated by LASSO without data filter

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Intercept</th>
<th>VB</th>
<th>( f )</th>
<th>( r_x )</th>
<th>( f )</th>
<th>( QT )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_x )</td>
<td>1.51</td>
<td>5.62</td>
<td>27.7</td>
<td>-5.53</td>
<td>-4.55·10^{-3}</td>
<td></td>
</tr>
<tr>
<td>( f )</td>
<td>1.01</td>
<td>1.13·10^{-3}</td>
<td>4.61·10^{-2}</td>
<td>-0.480</td>
<td>-62.5</td>
<td></td>
</tr>
<tr>
<td>( QT )</td>
<td>cool</td>
<td>QT</td>
<td>VB</td>
<td>( v_c )</td>
<td>( t_c )</td>
<td>( a_p )</td>
</tr>
<tr>
<td></td>
<td>-1.49·10^{-3}</td>
<td>-4.66·10^{-3}</td>
<td>-6.88·10^{-6}</td>
<td>0.839</td>
<td>-2.42</td>
<td></td>
</tr>
</tbody>
</table>

The majority of monomials selected by LASSO include process parameters which are classified as significant for the explanation of the roughness in Figure 1. Still the model is difficult to interpret on the basis of the coefficients. Therefore, the sensitivity of \( R_z \) was evaluated in the central points of the input parameters intervals and is visualized in Figure 2 a)-g). In each diagram, the respective process parameter was varied while the remaining model inputs were kept constant. The feed has the highest impact on \( R_z \) followed by \( VB \) and \( r_x \). One must note that the evaluated central points often do not meet process parameters used in the tests. For process cooling, other values than 0 or 1 are physically not even meaningful. Under these conditions, the presented absence of implausible nonlinearities is a sign of a robust model without overfitting. In Figure 2 h), the good agreement of measured and predicted \( R_z \) values is depicted, which suits the model quality reported in Table 1. Figure 2 i) proves the poor agreement of the measured \( R_z \) values and those calculated by the kinematic equation (1), which emphasizes the need for an empirically validated roughness model.

Figure 2. (a – g) Sensitivity of \( R_z \), LASSO model to the process parameters. Agreement of prediction for (h) LASSO and (i) kinematic roughness.
As mentioned before, the database was filtered selectively in order to fit less complex models by LASSO. When only dry cutting is considered, the database has 66 observations instead of 77 and the respective model 12 DOFs, RMSE=1.65 μm and R²=0.86. Further neglecting the tool corner radius of 1.2 mm still incorporates 50 observations. In this case, only 6 DOFs are needed for the model quality given in Table 2. This indicates that cooling and the tool corner radius have a strongly nonlinear impact. In such cases, splitting up models by filtering the database is helpful. The thus identified model coefficients are listed in Table 4.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Intercept</th>
<th>f</th>
<th>VB</th>
<th>QT</th>
<th>a_0</th>
<th>a_9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>0.977</td>
<td>28.0</td>
<td>-0.031</td>
<td>0.038</td>
<td>-30.0</td>
<td>-2.39</td>
</tr>
</tbody>
</table>

### 4.2. Modeling of the hardness increase

Before modeling the hardness increase, the database is analyzed by the help of a pairwise plot matrix, which is depicted in Figure 3. The weak correlations of the inputs indicate a fairly balanced dataset. The correlations of the hardness with the feed and the tool wear are physically reasonable. Both inputs increase mechanical surface loads and are thus driving forces of work hardening and grain refinement [1]. The cutting velocity primary increases thermal loads, which are known to fortify grain refinement. This could explain the present correlation with the hardness increase. The positive correlation with the tempering QT indicates that an originally softer material has a higher potential for mechanically induced hardening.

The diagrams show that the tool wear has a high impact on surface modifications, which is in line with findings in [2]. While the remaining sensitivities seem moderate, the bilinear model structure must be noticed. E.g. a quadratic hardness increase generated by higher feed and tempering is not represented in the diagrams. When such a model is used in a process control, the parameter sensitivities in the present working point can be accomplished by partial differentiation of the model equation. The agreement of measured and predicted hardness increases is depicted in Figure 4 e). Taking into account the inevitable deviations of hardness testing due to material inhomogeneity, which were analysed in [4], the model

![Figure 3. Pairwise correlation matrix for the hardness dataset](image-url)
quality is satisfying. When physical interpretability was less important than model quality, support vector machine or neural network algorithms could be applied as well.

Additionally considering specimens machined with process cooling increases the database from 47 to 51 observations. In this case, LASSO determines 8 instead of 5 DOFs and reaches RMSE=31.65 HV and R²=0.72. This confirms the presumed nonlinear impact and motivates further experiments in order to identify a separate model for turning with process cooling.

5. Conclusion

Machining of hardened steels can cause a variety of surface modifications, which were often analyzed qualitatively [1-4]. The deliberate adjustment and control of beneficial surface states by machining parameters however requires quantitative models. In this paper, the data driven generation of analytical models for the Vickers hardness increase and the average maximum profile height Rz was presented. The presented parameter sensitivities are physically plausible and in line with previous publications. Since Rz was closely correlated to the mean profile height Ra, it was decided to forgo an additional modeling of Ra. The efforts for the conduction of machining tests and subsequent surface analyses often lead to rather small datasets with less than 100 observations, while a large number of input parameters must be regarded, e.g. up to seven in this work. Higher order polynomial model approaches aggravate the problems of collinear input parameters and overfitting. As a result, the cross validated model performance measures deteriorated with additional DOFs, as shown for the fully quadratic approach. It was further shown that the parameter selection algorithm SWR does not solve this problem. LASSO however leads to good model qualities while it effectively reduces the number of DOFs. In subsequent works, the algorithm will be tested for the fitting of residual stress characteristics. Stress measurements are expensive and thus usually available in a low number, yet the dependencies with process parameters must be generally taken as nonlinear. The Elastic Net algorithm may generate models with even better quality measures, but this usually comes with a larger number of DOFs. In future applications, the generated models shall be tested for the turning of surfaces with a well-defined roughness and hardness increase. This requires the knowledge of the present tool and material state. A change of the tool wear may further require parameter adjustments in order to still meet the requirements. For this task, the model sensitivities are helpful, which simply result from partial differentiation of the identified equations. Consequently, the presented cross validated surface state models are a prerequisite for the controlled turning of AISI 4140 QT, which is the overall goal of the presented work.

Acknowledgements

The scientific work has been supported by the DFG within the research priority program SPP 2086 (SCHU 1010/65-1, LA 2351/46-1, WO 903/4-1). The authors thank the DFG for this funding and intensive technical support.

References