Clustering of Learning Sub-models for Quality Prediction in a Resource-Efficient Tool Grinding Process

Berend Denkena, Marcel Wichmann, and Michael Wulf(✉)

Institute for Production Engineering and Machine Tools Hannover,
An der Universität 2, 30823 Garbsen, Germany
wulf_m@ifw.uni-hannover.de

Abstract. The prediction of workpiece quality in process planning, using machine learning models, is a common-researched topic. Until now, trained models were static and could not update themselves with new data. However, this aspect is crucial when considering the continuously changing manufacturing circumstances in regards to new process parameters, materials, and workpiece geometries. In addition, repeatedly training process models with an extended mixed dataset decreases the prediction quality due to the increased data divergence. This paper presents an approach to automatically generate sub-models, which maintain the prediction quality even if novel data is considered. The challenge is to define the amount and content of these sub-models through clustering. Tool grinding experiments will be conducted with different process parameters, materials, and workpiece geometries in order to obtain a divergent dataset. Subsequently, cluster approaches are compared to obtain dynamic growing models, which enable optimized planning for a more resource efficient process. Finally, the method will be generalized in order to ensure a process-independent usage.

Keywords: Machine Learning · Process Models · Tool Grinding

1 Introduction

Besides empirical and numerical modelling approaches, machine learning methods are being used more and more to describe relationships in a machining context [BRI06, DEN19, KÖN17]. The future vision of an autonomous machine tool needs to include this knowledge in order to perform a successful production [DIT21]. Because of their ability to map even complex and comprehensive processes, the automatic generation and evaluation of machine learning models is frequently researched [DEN20, KRÜ19]. In this research, the limits of modelling are specifically focussed on due to the limited complexity a model can include. To visualize this limit, Fig. 1 depicts three synthetically generated datasets and the support vector machines (SVM) of various dataset combinations. Their quality is evaluated using the mean absolute percentage error (MAPE) which is defined as the absolute model error per mean value of the modelled target parameter. A perfect model has a MAPE of 0%, while a larger MAPE indicates lower

T. Bauernhansl et al. (Eds.): WGP 2023, LNPE, pp. 94–103, 2024.
https://doi.org/10.1007/978-3-031-47394-4_10

model qualities [MYT16]. The combination of Dataset 1 and 2 in Fig. 1 results in a model quality between them. The 5.8%-MAPE of Dataset 2 is nearly average with the higher MAPE of Dataset 1 because all datasets have the same size. This effect can't be observed for the combination of all three datasets. Due to the inauspicious location of Dataset 3, the MAPE is even greater when compared to the single datasets' MAPEs. A possible solution could be clustering the datasets into two data clusters with one model each, instead of one model including all of the datasets. One cluster with Dataset 3 and another including Dataset 1 and 2 yields a compromise with fewer models and higher model quality, despite the included data complexity.

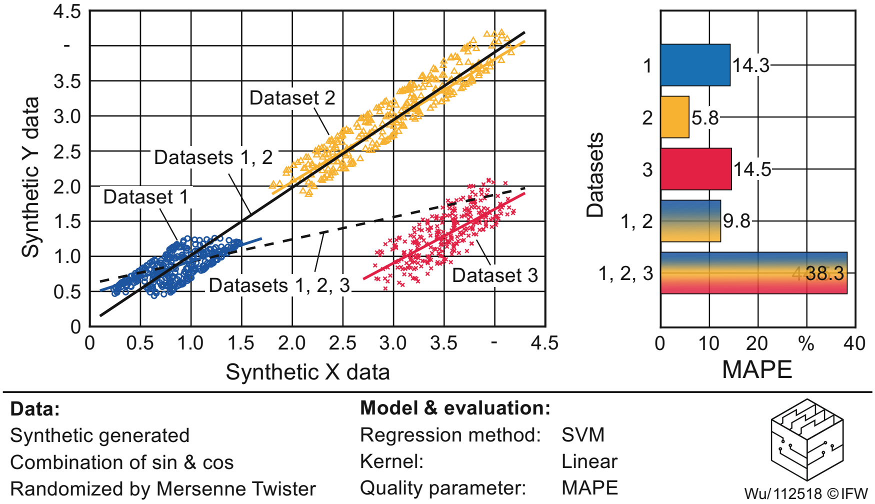

Fig. 1. Limited complexity of linear support vector machines on synthetic data

Currently, it is common practice to generate a separate model for each dataset (e.g. for each material-tool combination), resulting in a large amount of different models [DIT20, UHL21]. This process means that any new data, e. g. of a slightly different material, cannot be processed because the models have only a small definition range. Thus, an approach by which models can adapt to the database, and thus process new data points, is missing. This capability is essential for an autonomous machine tool with self-optimizing process planning and is, therefore, the subject of this paper.

In contrast to machine learning approaches, there is only limited research on clustering approaches in a machining context. Ochel et al. cluster continuous milling data as current signals, as well as axis position signals, to differentiate between geometrical workpiece features [OCH22]. Before the similarity to the predefined signal patterns per feature is calculated, the signals are roughly analysed by the spindle speed to omit tool changes. This approach is used by Brecher et al. to create a tool wear model for milling processes [BRE22]. It becomes clear that not just for the previously described data processing, but even for machine learning and clustering purposes, process knowledge is crucial [BRE20]. Those self-optimizing systems need to be able to automatically adapt to changing inputs [MÖH20]. For this aim, manually clustering using expert knowledge must be structured and automated by algorithm-based methods.

2 Approach

Based on the literature, before am automated hierarchical cluster algorithm is discussed as step towards an autonomous machine tool, manually clustering according to the most dominant process input parameters is presented. Analyzing both strategies successively, makes it possible to develop a general method of clustering for incorporating new data while maintaining prediction quality. The used dataset originates from helical flute grinding experiments conducted during the manufacturing of shank tools. Detailed information about the data is listed below.

- Machine tool: Vision 400 L (Walter Maschinenbau GmbH), Fanuc 31i-B5 control.
- Workpiece: Diameter D = 6, 10, 12 mm and core diameter D_c = (0.3, 0.5) x D; cemented carbide from Tigra GmbH with amount of cobalt = 3, 6, 10%.
- Grinding tool: Hybrid bonded diamond grinding wheels, shape 1A1, grain size = D33, D46, D54 μm.
- Process parameters: Feed rate v_f = 50, 100, 200 mm/min; cutting speed v_c = 15, 18, 22 m/s. Between two and three executions to ensure statistical reliability.
- Simulation data: Local cutting conditions by the software IFW CutS® in terms of local contact length l_g and equivalent chip thickness h_{eq} as established in [DEI10].
- Target parameter: Spindle current in % adjusted for frictional effects named DTRQ.

There are many methods for evaluating the found clusters [FRA10]. The given use case is to model the spindle current value DTRQ of helical flute grinding processes. Here, clustering should lead to more optimized regression model qualities within a set of clusters when compared to a comprehensive single model. The quality of a single cluster is evaluated by generating a regression model. The average quality of a cluster set can then be used for comparison with other sets. The quality of a single model can be expressed as a combination of the MAPE and the determination coefficient R^2. The latter describes the variance of explanation for due to the model in a range from 0 (not explained) to 1 (fully explained) [BAC16]. The equal modelling method is used for the regression models of all clusters in every cluster set, to ensure comparability. A linear SVM was chosen as the modelling method because of its short calculation times and successful performance in preliminary investigations [UHL22]. The modelling is done by Mathworks® Matlab using standard hyper parameters since an optimization of them wouldn't improve the model quality. Reliable regression models have similar prediction qualities in regards to both the validation data and the test data. Otherwise, the model is said to overfit or underfit. To omit models with an indication of under- or overfit, the MAPE and the R^2 of validation and testing were compared using a fitting indicator *Fit*, as established in Eq. 1. Preliminary work urge the use of models with $Fit \leq 10$. To ensure a statistical certainty, the modelling was done at least 20 times, for every model, with a 5-fold-cross validation and 20% unknown test data.

$$Fit = 0.5*\left(\frac{MAPE(Testing)}{MAPE(Validation)} + \frac{R^2(Validation)}{R^2(Testing)}\right) \tag{1}$$

3 Manually Clustering Using Expert Knowledge

Defining clusters based on expert knowledge, strongly depends on the given use case. In terms of tool grinding, the available process input parameters, presented in the previous section, are possible indicators for clustering. Figure 2 presents the model quality for DTRQ of the grinding wheel D46 engaging a workpiece with 10% cobalt (D46T10). In the upper graph, with an increasing amount of D46T10 data points, a mostly constant model quality is already achieved using 30 data points. Including 60 data points, the final modelling quality is reached (green line in Fig. 2). This result shows that at least 30 observations are necessary in order to obtain a SVM regression model for helical flute grinding processes with an acceptable model quality as also found in [DIT20].

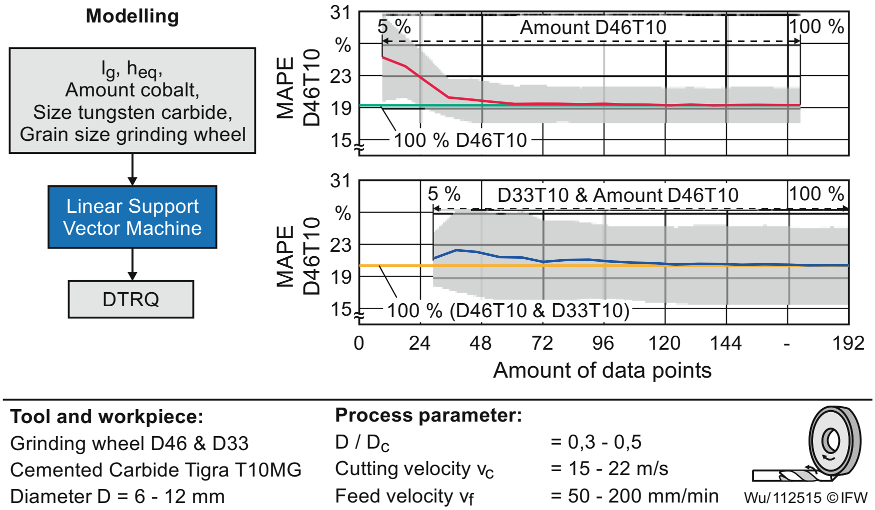

Fig. 2. Differing data influencing the quality of a common regression model (Color figure online)

For the lower graph of Fig. 2, additional data in terms of D33T10, is used to train the model. In comparison to D46T10 the datasets differ in the grinding wheel's grain size, which is reduced from 46 μm to 33 μm. Due to this variance, there is a base of 30 data points, which is consecutively increased by the D46T10 data. Even with this additional 30 data points of D33T10, when compared to the upper graph, the final model quality of D46T10 values is first reached with about 110 data points (30 out of D33T10 and 80 out of D46T10) in terms of the mean value. The modelling complexity in the lower graph was too high, thereby preventing the final model quality from being achieved before the 110th data point. In addition to the mean value, model variance is also affected by additional D33T10 data. This additional data allows for an even earlier attainment of the final model quality with 5% of D46T10 (lower graph) as compared to 9.5% of D46T10 (upper graph). Furthermore, the resulting range of variation for the MAPE doubles independently of the ratio of the data points from D33T10 to D46T10 included in the training data. This shows a crucial influence of the training datasets' composition on the model quality.

Besides the grinding wheel specification, other process variables are also possible for clustering based on expert knowledge. Surprisingly, models created with data that was clustered after the workpiece geometry parameters D or D_c fail immediately. In contrast, clustering by v_f or using local cutting conditions as a cluster indicator yields reliable performance as presented in Fig. 3. It becomes clear, that a combined dataset, including all three stages of feed rate, performs up to 7.7% (MAPE) worse than the individual stages' models. This difference is visible in both the MAPE and the R^2 model quality criterion. Remarkably the MAPE for the $v_f = 200$ mm/min dataset is clearly lower when compared with the $v_f = 100$ mm/min dataset, while their R^2 is mostly the same. This result confirms the differences in model quality, at least in regards to the used quality measurement method, are thereby supporting the use a combined MAPE and R^2 for objective model evaluation.

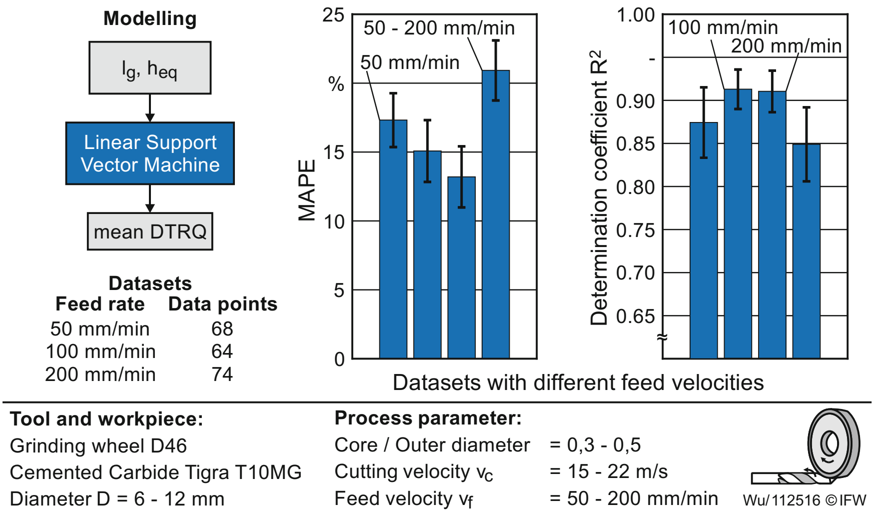

Fig. 3. Model quality after clustering D46T10 by feed rate v_f

The presented results lead to the following boundary conditions for generating and evaluating clusters for regression models:

1. The approach shall generate clusters with at least 30 data points each to ensure a proper model quality (in terms of helical flute grinding).
2. Additional data, which differs from the existing data, can drastically decrease model quality. The approach shall be able to identify and split those datasets into clusters.
3. Even a small number of additional data added to a huge dataset are able to reduce the model quality. The approach shall be sensible for those data points and allocate them into a new cluster.
4. Considering the previous boundary conditions, the approach shall minimize the number of clusters (and with this, the number of models).
5. The generated models should be evaluated based on several quality metrics.

The major disadvantage of this knowledge-based clustering is that well-educated and experienced employees are necessary to cluster the data. They have to decide which process parameter is sufficient as a cluster indicator, or even, if a combination of several parameters has to be taken into account. In addition, they have to define which data could decrease the model quality when combined with other data, although there is no established approach for measuring this data divergence. To conclude, besides the process knowledge, brief insights into data preprocessing and model creation are required for knowledge-based clustering. Another approach is to use hierarchical cluster algorithms which can handle the confusing number of data points systematically and within shorter durations.

4 Clustering Automated by a Hierarchical Algorithm

Hierarchical methods are well researched and often used in combination with a euclidean distance criterion to split data into clusters [FRA10]. The hierarchical method starts with a cluster for each point of the dataset and calculates a matrix of differences between every possible pair of clusters. Subsequently, the two clusters with the shortest distance are merged into a common cluster, whereby the number of clusters for the whole dataset decreases. The repetition of this procedure with new distance matrices and merged clusters leads, in the end, to one single cluster containing every point of the dataset. The generated structure of clusters is called "dendrogram" and enables the classification of data points into a given number of clusters [FRA10].

Differences in this hierarchical cluster method can be seen in the cluster linkage for euclidean distance measurements. For example, the distance between two clusters can be calculated by the shortest distance between two points of the clusters or by their farthest distance. There are many different linkage criteria, but to obtain clusters with low variance and comparable sizes (see Sect. 2), the ward linkage method is more suitable [FRA10, WAR63]. The method aims to minimize the overall error sum of squares (ESS) of a dataset with c points $P_{ij} \in \mathbb{R}^n$ in a cluster j by calculating ESS for every possible merge of two clusters, as presented in Eq. 2.

$$ESS = \min \sum\nolimits_j^d \sum\nolimits_i^c \left(P_{i,j} - O\left(\mathbb{R}^2\right)\right)^2 \tag{2}$$

The model quality of hierarchical clustered datasets with the ward linkage is depicted in Fig. 4 for different clustering input parameters and numbers of clusters. The grey bars going in the vertical direction present the model quality in regards to the whole D46T10 dataset as one single cluster. The horizontal bars in blue and yellow depict two, or rather, three clusters out of the D46T10 dataset. These clusters were generated by the cluster algorithm using different input parameters. As a benchmark, the manual clustering that utilized the $\mathrm{v_f}$ presented in Fig. 3 is done with this approach too.

Figure 4 points out that a clustering by the local cutting conditions $\mathrm{l_g}$ and $\mathrm{h_{eq}}$ is also suitable for achieving an improved model quality in comparison to the overall model with only one cluster (grey bar). In contrast to manual clustering, this method combines at least two input parameters to create the set of clusters. Because of the numeric optimization seen in Eq. 2, a manual approach can't achieve this same clustering in manageable time periods. Across all input parameters, Fig. 4 shows that three clusters (yellow) tend to result in in an average 2.7% lower MAPE than the two clusters (blue). However, with an increasing number of clusters, the Fit indicator increases with a mean of 8.2%, thereby decreasing model quality. This result supports the boundary conditions of Sect. 2, in regards to prefer a minimal amount of models.

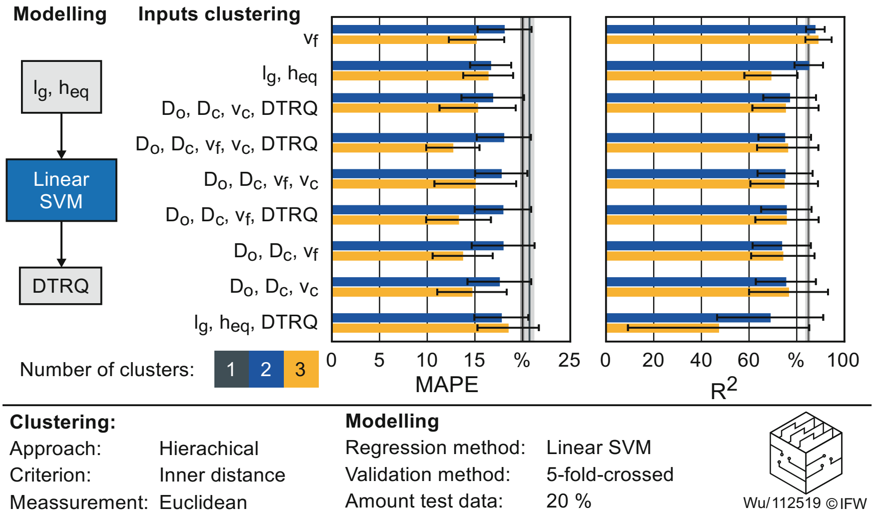

Fig. 4. Model quality of hierarchical clustered datasets by different clustering inputs (Color figure online)

The benefit of this automated approach, in comparison to the expert knowledge needed for manual clustering, is the structured proceeding that leads to an optimized clustering. The user can focus more on the evaluation and comparison of clusters rather than generating and testing them. Even combinations of process input parameters can be used as cluster inputs in this approach, which is more challenging and time consuming in the manual way. As presented in Fig. 4, an individual cluster isn't necessary for every stage of a process parameter. The algorithm-based approach can test every combination of process parameters for every number of desired clusters. This leads to reliable process models that enable more resource efficient production. The following section sumarizes the approach and gives guidance on how to modify the inputs for clustering purposes.

5 Generalization of the Method

A general description, as presented in Fig. 5, illustrates the usage of the automated clustering approach with a hierarchical algorithm independently of the machining process. On the left side, the regular use of the process model is depicted with unlabelled data and the predictions are calculated with the current model. The latter can be used, for example, in process planning or quality prognosis. When newly labelled data becomes available, the current model's quality can be evaluated by predicting the new input data.

Therefore, the current model is used to generate the labels for the data. A comparison of the predicted labels with the actual ones gives the prediction quality of the current model for the new, labelled data. MAPE, R^2, and Fit are used to ensure a sustainable assessment of model quality. If this new prediction quality equals or exceeds the current prediction quality, no action is needed. Conversely, new clusters and corresponding new models would have to be generated. In order to account for influences from measurement uncertainties and model building on the forecast quality, a threshold can be introduced. For example, clustering would only be triggered at a 5% reduction in model quality. To account for any possible influence on model quality, a new clustering is also conceivable for the first variation of a previously constant process input variable.

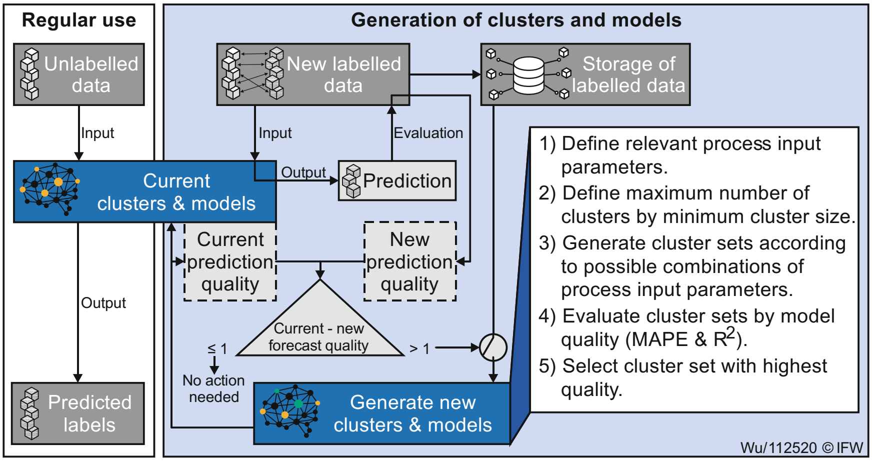

Fig. 5. Method of clustering for incorporating new data while maintaining prediction quality.

Generating new clusters requires the definition of relevant process input parameters included in the dataset. For example, helical flute grinding would include feed rate, cutting speed, local cutting conditions, workpiece material, and grinding wheel specifications. The maximum number of clusters is determined by the minimum number of data points per cluster. For every combination of the defined process input parameters a set of clusters is generated with the automated hierarchical cluster algorithm. This process enables the user to evaluate each cluster set. Therefore, a regression model is built for every cluster to compare the model quality. The most suitable set of clusters and models is then chosen to become the new current set for regular usage.

6 Conclusion and Outlook

The paper presented the limits of machine learning models in the case of complex tool grinding datasets concerning their dynamic growth during production. To overcome this challenge, datasets can be clustered to generate a set of models, instead of just a single one, for modelling the process relations. This clustering requires knowledge about the number and subdivision of the clusters. Both a manual and an algorithm-based approach for splitting datasets into suitable clusters were analysed to develop a general method of automated clustering. This method allows more than one process parameter to be considered in clustering, resulting in a reduced number of clusters, creating high quality process models for even complex datasets. Reliable process models are required in process planning to ensure a resource-efficient production. It can be seen as a step towards self-optimizing process models that automatically adapt to new processes. Such systems are necessary in order to achieve the vision of an autonomous and energy-efficient machine tool. Further work is necessary to explore this developed approach, including different linkage criteria and algorithms, like a density-based method. The presented approach uses a linear SVM to model the process relations and evaluate the clusters. Other methods, such as neuronal networks or gaussian process regressions, should be investigated too. Additionally an automated selection of the best set of clusters as well as the required calculation effort should be explored too.

Acknowledgement. The authors would like to thank the German Research Foundation (DFG) for funding the project LearnWZS - Learning process adaptation for tool grinding (number 445811009), which enables this investigation. Furthermore, the authors thank the Sieglinde-Vollmer Foundation for supporting this research.

References

Backhaus, K., Erichson, B., Plinke, W., Weiber, R.: Multivariate Analysemethoden (14. Aufl.). Springer, Berlin, Heidelberg (2016). https://doi.org/10.1007/978-3-662-46076-4

Brecher, C., Ochel, J., Lohrmann, V., Fey, M.: Maschinelles Lernen zur Prädiktion der Bauteilqualität. Zeitschrift für wirtschaftlichen Fabrikbetrieb **115**(11), 834–837 (2020)

Brecher, C., Lohrmann, V., Wiesch, M., Fey, M.: Clustering zur Bestimmung von Werkzeugverschleiß. Zeitschrift für wirtschaftlichen Fabrikbetrieb **117**(4), 218–223 (2022)

Brinksmeier, E., et al.: Advances in modeling and simulation of grinding processes. CIRP Ann. **55**(2), 667–696 (2006)

Deichmueller, M., et al.: Determination of static and dynamic deflections in tool grinding using a dexel-based material removal simulation. In: CIRP 2nd International Conference Process Machine Interactions 2010. Vancouver, Canada (2010)

Denkena, B., Dittrich, M.-A., Böß, V., Wichmann, M., Friebe, S.: Self-optimizing process planning for helical flute grinding. Prod. Eng. Res. Devel. **13**(5), 599–606 (2019)

Denkena, B., Dittrich, M.-A., Lindauer, M., Mainka, J., Stürenburg, L.: Using AutoML to optimize shape error prediction in milling processes. MIC Procedia **20**(1), 160–165 (2020)

Dittrich, M.-A., Uhlich, F.: Self-optimizing compensation of surface deviations in 5-axis ball-end milling based on an enhanced description of cutting conditions. CIRP J. Manuf. Sci. Technol. **31**(1), 224–232 (2020)

Dittrich, M.-A.: Autonome Werkzeugmaschinen - Definition, Elemente und technische Integration. Habilitation, Gottfried Wilhelm Leibniz Universität Hannover (2021)

Frades, I., Matthiesen, R.: Overview on techniques in cluster analysis. In: Bioinformatics Methods in Clinical Research, pp. 81–107. Humana Press, Totowa, USA (2010)

Königs, M., Wellmann, F., Wiesch, M., Epple, A., Brecher, C.: A scalable, hybrid learning approach to process-parallel estimation of cutting forces in milling applications. In: WGP-Jahreskongress Aachen 2017, vol. 7, pp. 425–432. Apprimus, Aachen (2017)

Krüger, J., Fleischer, J., Franke, J., Groche, P.: WGP-Standpunkt KI in der Produktion. Wissenschaftliche Gesellschaft für Produktionstechnik WGP e.V. (2019)

Möhring, H.-C., Wiederkehr, P., Erkorkmaz, K., Kakinuma, Y.: Self-optimizing machining systems. CIRP Ann. **69**(2), 740–763 (2020)

Myttenaere, A.D., Golden, B., Le Grand, B., Rossi, F.: Mean Absolute Percentage Error for regression models. Neurocomputing **192**, 38–48 (2016)

Ochel, J., Fey, M., Brecher, C.: Semantically meaningful segmentation of milling process data. In: Behrens, B.A., Brosius, A., Drossel, W.G., Hintze, W., Ihlenfeldt, S., Nyhuis, P. (eds.) Production at the Leading Edge of Technology, pp. 319–327. WGP 2021. LNPE. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-78424-9_36

Uhlich, F.: Lernende Prozesssimulation für die Prognose und Kompensaiton von Formabweichungen in der Einzelteilfertigung. Dr.-Ing. Diss., Gottfried Wilhelm Leibniz Universität Hannover (2022)

Ward, J.H.: Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. **58**(301), 236–244 (1963)