Application of grey-box modeling for machine state prediction in manufacturing

Utsav Awasthi[a] and George M. Bollas [a,1]

[a] Department of Chemical & Biomolecular Engineering, UTC Institute for Advanced Systems Engineering, University of Connecticut, 159 Discovery Dr, Storrs, CT, 06269, USA.

Abstract

The recent technological advancements and shift towards digitalization in the manufacturing industry have led to the development of advanced methods to build data-driven models. The development of data-driven models has been made possible because of the improved data collection capabilities and advanced model-building methodologies. However, these models are limited in their application to the process for which they have been developed. Also, the data-driven models are black-box and provide no insight into the underlying model structure. To overcome these challenges, hybrid modeling methods that build grey-box models are being used increasingly because of their ability to develop models that are dependent on the data of the process and also based on the physics of the process. This study explores a hybrid model-building methods based on symbolic regression to build physics-based surrogate models. This hybrid modeling method is then demonstrated by building a physics-informed surrogate model for a complex physical phenomenon of precision machining tool wear.

Keywords

Grey box model, surrogate modeling, symbolic regression

Introduction

With the advent of the fourth industrial revolution, the manufacturing industry has developed enhanced technology to generate high-quality data. This, coupled with the improved computing capabilities to use data, has helped develop data-driven models. Building a data-driven model for a process requires good quality data that captures the behavior of the entire physical process. Data-driven models utilize pattern recognition algorithms to build classification or regression models of a physical phenomenon. These models help capture the behavior of unknown, complex factors that influence a process, which cannot be captured using the first-principle/mechanistic models. This attribute of data-driven models can help first principle models improve their predictions. These data-driven models are good at capturing a phenomenon, but these models are limited to the data that are used to build the model. Furthermore, there is no insight into the model as these models are black-box. To overcome this challenge, many recent studies are exploring hybrid models. Hybrid models are developed using data and the knowledge of the physics of the process. Sansana et al. (2021) categorized hybrid models based on the structure of the model. A hybrid model has a serial structure when the data-driven approach is used to model a phenomenon that is too complex to model using a first principle model. A parallel structure is utilized when some effects of the process cannot be captured by a first principle model. In this case, a parallel structure helps to improve the predictions. Both serial and parallel hybrid model structures improve the model prediction of the first principle models. The third category of hybrid models is surrogate models or grey-box models. Surrogate models are a simple approximate representation of a complex process in the form of a mathematical model developed using minimal data. In the manufacturing industry, surrogate models are being used increasingly because these models are simple to develop and are computationally less expensive than first principle or data-driven models. They can be used for process monitoring, optimization, and predictive maintenance. Here, we discuss a few popular grey-box model/surrogate model building toolboxes that can be used to build physics-informed models. Two of these toolboxes, namely ALAMO (Automated Learning of Algebraic Models for optimization) and GPTIPS (Genetic Programming Toolbox for Identification of Physical Systems), are used to build generic surrogate models of the fault that are physics-informed, in a manufacturing process.

Surrogate modeling methods

Surrogate models are approximate models of a physical process cast as simple mathematical expressions. Several software toolboxes can be used to develop surrogate models with minimal data. A few of the modeling software that have been used to build surrogate models are Eureqa

[1] Corresponding author: G. M. Bollas
(E-mail: george.bollas@uconn.edu).

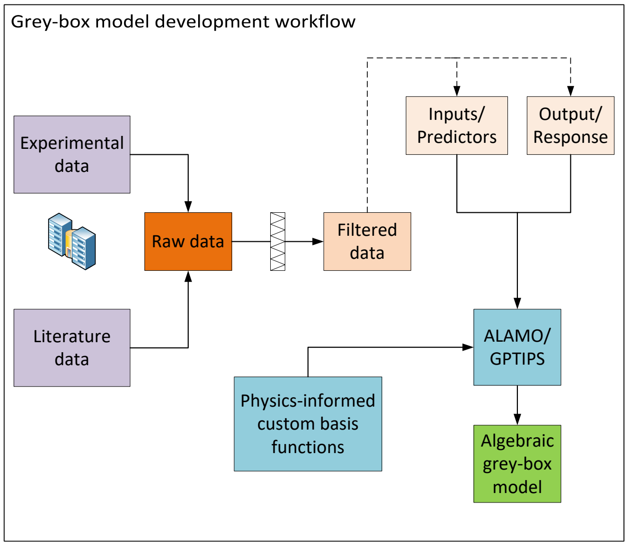

Figure 1: Workflow of tool wear surrogate model development.

(Schmidt and Lipson, 2009) (now DataRobot), AI Feynman: A physics-inspired method for symbolic regression (Udrescu and Tegmark, 2020), Automated Learning of Algebraic Models for optimization (ALAMO) (Cozad et al., 2014, 2015; Wilson and Sahinidis, 2017), and Genetic Programming Toolbox for Identification of Physical Systems (GPTIPS) (Searson et al., 2010; Searson, 2015). These toolboxes apply symbolic regression to develop a model in the form of an algebraic expression. Symbolic regression is a kind of regression analysis that gives a mathematical expression that best fits the design space. This approach is different from the traditional regression methods because, in symbolic regression the model structure and the regression coefficients are determined simultaneously. In developing a surrogate model, first, the input features that influence the response variable are selected, along with proper relevant basis functions. These basis functions vary from software to software, but primarily consist of simple mathematical operators, such as $+$, $-$, $\backslash$, $\times$, $\surd$, exp, log, ln, sin, cos, tan, x^2, x^3, etc. Many software programs have the capability to provide user-defined custom basis functions that can be used to infuse domain expert knowledge in building the model. These custom basis functions are simple, functional forms of the inputs used to build the model. As mentioned earlier, these custom basis functions can be used to provide expressions that have actual physical significance, and can be used to infuse the information of the physics of the process and assist the toolboxes in coming up with mathematical models for the response variable that has a physical meaning. The best fit model is selected by minimizing a metric corresponding to the overall model prediction error with the actual values of the response variable. To obtain a model of a good fit, these software packages have several hyperparameters that can be tuned, besides selecting the mathematical operators or the basis functions. For example, in GPTIPS, the number of genes, depth of the tree, and mutation depth are some of the hyperparameters that can be tuned for model building. In this study, we discuss how to build physics-informed surrogate models or grey-box models using ALAMO and GPTIPS as the software toolboxes. These software packages are discussed in more detail below.

Automatic Learning of Algebraic models (ALAMO)

The Automatic Learning of Algebraic models software (ALAMO) generates algebraic models using machine learning and optimization methods (Cozad et al., 2014; Wilson and Sahinidis, 2017). To develop a low-complexity surrogate model using minimal experimental data, ALAMO follows a series of steps. First, an evenly spaced initial sample space is defined as the training set to build a surrogate model. In the second step, integer optimization is used to select the basis functions to build a model for the initial data set. In the third step, adaptive sampling is performed using derivative-free optimization methods to find a data point for which the model does not predict accurately. This data point is added to the training set, and steps 2 and 3 are repeated to obtain a model for which the model accuracy does not improve any further. ALAMO provides 8 fitness metrics to select for model building. These are the Bayesian information criterion (BIC), Mallow's Cp (Cp), correlated Akaike's information criterion (AICc), Hannan-Quinn information criterion (HQC), mean square error (MSE), the sum of squared error plus a penalty proportional to the model size, risk information criterion (RIC), maximum absolute deviation plus a penalty proportional to model size (MADp). In addition, the user can select between the basis functions of, exp, ln, log, sin, cos, and constant. To build a model, values of the basis functions that are power of a term (Monomial power), power of the product of two terms (Multi2Power), power of the product of three terms (Multi3Power), power of the ratio of two terms (RatioPower), can be tuned to improve the model fit. Further, user-defined custom functions can be provided to ALAMO. These models are functions of the inputs and are considered by ALAMO during model building. Physics-based functions can be provided as custom basis functions to make physics-inspired surrogate models. More details on ALAMO can be found in Cozad et al. (2014, 2015) and Wilson and Sahinidis (2017).

Genetic programming toolbox for Identification of Physical systems (GPTIPS)

The Genetic programming toolbox for Identification of Physical systems (GPTIPS) is a free, genetic programming-based toolbox to build models using symbolic regression. Genetic programming (Koza, 1994) is a technique to evolve programs by performing operations similar to the biological operation performed by genes, such as mutation, crossover, etc. GPTIPS performs symbolic regression to build models in simple algebraic forms of function of the input/predictor variables. GPTIPS generates equations using multigene genetic programming, where the response variable is weighted functions of genes (represented in the form of trees). These genes (tree structures) are evolved iteratively by performing generation, mutation, and crossover. GPTIPS provides a variety of tournament selection approaches to choose from. These tournament selection options are regular tournament selection, Pareto tournament selection, and lexicographic tournament section. GPTIPS also has a large set of basis functions $+$, $-$, $\times$, $\backslash$, mult3(product of three terms), add3(sum of three terms), tanh, cos, sin, exp, $\log_{10}$, x^2, abs, x^3, $\sqrt{}$, $\exp(-x)$, if-then-else, $>$, $<$, $\exp(x^2)$), threshold and step functions, to select from for model building. Similar to ALAMO, user-defined functions can also be provided to GPTIPS. Because, GPTIPS is an open-source toolbox, where the fitness criterion for the model building can be defined by the user. The default fitness metric used by GPTIPS is the root mean squared error (RMSE) value. GPTIPS has an extensive list of settings or parameters that linked to the genetic programming that can be selected and tuned for model building. Few of the settings are population size, number of generations, maximum tree depth, maximum mutate depth. More details on GPTIPS and the model building parameters and setting options can be found in Searson et al. (2010) and Searson (2015).

Application of surrogate modeling methods in a manufacturing process

Computer numerical control (CNC) machines are the backbone of the manufacturing sector, which spans from small-scale use to manufacturing the parts for the aerospace and automobile industries. Milling is one of the operations performed in these industries using CNC machines. During milling, a circular tool with high-grade metal inserts is used to remove material from a workpiece in the form of metal chips, which results in high shear stresses, friction between tool and workpiece, and high temperature at the interface. These extreme conditions cause the abrasion of the inserts on the tool, resulting in the tool to get worn out, which impacts the surface finish of the workpiece. If the tool is not replaced, it may damage the workpiece and the machine. Admissible machine settings include the spindle speed, feed rate, width of cut, and depth of cut. These settings control the cutting operations and can impact/wear the tool. The properties of the workpiece and the tool will also influence the tool wear. The machine settings need to be considered in a model of tool wear. In this study, the same workpiece material and tool were used for data generation. Therefore, the impact of the workpiece and material properties cannot be demonstrated due to lack of diverse data.

Machining data

Machining data were collected by experiments in a HAAS Mini Mill and a DMG MORI LASERTEC 65 3D hybrid machining centers. The HAAS mini mill is a 3-axis compact machining center, and the DMG machine is a hybrid machine that can perform milling as well as additive manufacturing. For both the machines, milling was performed on an AISI 4340 cylindrical steel block of 177.8 mm diameter and 20 HRC hardness using a tool with two Kennametal inserts of grade KC725M and a lead angle of 90°. Each of these steel blocks was called a "Part". A spiral tool path was selected to generate tool wear data and design a cylindrical boss of height 10.16 mm and a diameter of 76.2 mm. One complete spiral tool path was called a "Run", and while following the spiral tool path, the tool went around the workpiece for five times. The width of cut for this spiral path was 10.16 mm, and the depth of cut was 2.54 mm. At the end of the spiral tool path, when a boss of 76.2 mm diameter was obtained, the machining was stopped, and the flank wear measurement was taken. This whole process was repeated four times on a part to form a boss height of 10.16 mm and a diameter of 76.2 mm. This process was performed on the HAAS, and the DMG machines and tool wear data were collected. In the case of the HAAS machine, a total of 7 parts were machined, part#1-part#4 were machined for the machine settings of feed rate 710.184 mm/min and spindle speed 2330 RPM. The part#5-part#7 were machined for a feed rate of 970.483 mm/min and a spindle speed of 3184 RPM. For the experiments performed on the DMG machine, a total of 4 parts were machined with a feed rate of 970.483 mm/min and spindle speed of 3184 RPM. Table 1 shows the machine settings for the experiments. In the case of the HAAS machine, the flank wear measurements were estimated using the values of the tool radius change. The tool diameter was measured after each run, and for the last run of the part#7, the flank wear measurements were taken using a Keyence VHX-500, with a VH-Z100 lens and OP-72402 light ring microscope. To determine the flank wear, the tool radius change was calculated from the tool diameter measurements. In milling, the tool radius change follows the same trend as that of the flank wear. Therefore, by using the tool radius change values and the flank wear value for the last run, the flank wear values were estimated because they both were assumed to have the same profile. In the case of the DMG machine, the flank wear was measured using the Keyence VHX-500, with a VH-Z100 lens and OP-72402 light ring microscope. Table 2, shows the values of the machine settings of spindle speed, width of cut, depth of cut, and feed rate, that are used to build a surrogate model for tool wear using ALAMO and GPTIPS.

Custom basis functions

Symbolic regression was aided to come up with physics-informed, user-defined functions of the input/feature variables, as basis functions provided to ALAMO and GPTIPS. The custom basis functions provided were the mean material removal rate (*MRR*), mean material removed (*MR*), and the product of feed rate and cutting time.

The Mean material removal rate (*MRR*) is defined as the product of width of cut (a_e), depth of cut (a_p) and feed rate (f) (Awasthi and Bollas, 2020; Awasthi et al., 2022). *MRR* corresponds to the rate at which the material is removed from the workpiece, and high *MRR* values imply aggressive machining conditions. The second custom basis function was the mean material removed (*MR*), which corresponds to the amount of work done by the tool. The mean material removed was calculated as the product of the mean material removal rate with the cutting time (t_{cut}). The mean material removed is a good indicator of the condition of the tool because it considers the machine settings and the time for which

Table 1: Machine settings for the experiments performed on HAAS and DMG machines.

Machine	# of parts	Tool path	Diameter (mm)	Teeth	Width of cut (mm)	Depth of cut (mm)	Feed rate mm/min	Spindle speed RPM
HAAS machine	#1-#4	Spiral	20	2	10.16	2.54	710.184	2330
HAAS machine	#5-#7	Spiral	20	2	10.16	2.54	970.483	3184
DMG Machine	#1-#4	Spiral	20	2	10.16	2.54	970.483	3184

Table 2: Nomenclature of the symbols.

Symbol	Description
v_s	Spindle speed
f	Feed rate
a_e	Width of cut
a_p	Depth of cut
t_{cut}	Cutting time
s_i	Indicator
MR	Mean Material removed
MRR	Mean Material removal rate
W	Tool wear

Table 3: Custom basis functions for the surrogate models.

	Description	Expression
1	Material removal rate (MRR)	$a_e a_p f$
2	Material removed (MR)	$a_e a_p f t_{cut}$
3	Product of feed rate and time	$f t_{cut}$
4	Product of three inputs	$x_1 \times x_2 \times x_3$

machining was performed. The third custom basis function selected for the model building was the product of feed rate and cutting time. Feed rate is the machine setting that determines the speed at which the tool advances into the workpiece while cutting. Therefore, the product of the feed rate with the time of cutting would give us the distance the tool has moved while machining. The cutting distance is also an important quantity that would be correlated with tool wear. GPTIPS only accepts the functional form of an expression as a custom basis function. Therefore, the product of three inputs was used as the custom basis functions. These custom basis functions correspond to the expressions of mean material removed. The custom basis function used for model building are mentioned in Table 3.

Physics-informed surrogate model

To build the surrogate models, the machine settings selected as inputs were the spindle speed, v_s, width of cut, a_e, depth of cut, a_p, feed rate, f, and cutting time t_{cut}. In addition to the machine settings, an indicator feature, s_i, shown in Table 2, was also provided as an input. This indicator was a label representing the selection of the HAAS or the DMG data sets. The natural logarithm of the tool wear was taken as the output. The log transformation of the output was selected for model building because it reduces the variability of the data. All the input and output variables data were normalized for model building.

Table 4 shows the settings for ALAMO and Table 5 shows the settings for GPTIPS. In ALAMO, BIC was selected as the fitness metric to develop the tool wear model, and log, ln, and exp were selected as the basis functions. Monomial power, Multi2Power, Multi3Power were tuned to obtain a model of good fit. The detailed settings of model building are tabulated in Table 4. The Mean material removal rate, mean material removed, and product of feed rate and cutting time were provided as custom basis functions to ALAMO. To build a model using GPTIPS, $+$, log, exp, $\times$, x^3, $\sqrt{}$ were used as the basis functions. The product of three terms was provided as a custom basis function. In GPTIPS, the maximum number of genes per individual and the maximum depth of trees were the hyperparameters that were tuned to obtain a good model fit. To construct a GPTIPS model, the population size was selected to be 100, and the number of generations to run was selected to be 100. The optimal values of the maximum number of genes and the maximum tree depth obtained after tuning were 6 and 2, respectively, and are tabulated in Table 5. As shown in Table 6, both models had a high coefficient of determination value. The ALAMO model for tool wear had a R^2 value of 0.962, and the GPTIPS model had a R^2 value of 0.969. This shows that both models could capture the behavior of tool wear during machining and provide good predictions.

After tuning the hyperparameters, the best model obtained using ALAMO is shown in the first row of Table 6. The ALAMO model was a function of t_{cut}, $\exp(s_i)$, v_s^2, t_{cut}^2, t_{cut}^3, $v_s s_i t_{cut}$. As shown by Binder et al. (2017), the tool wear profile has three phases. The first is an initial increase in the tool wear, the second is a steady state wear, where the tool wear gradually increases, and the last is an exponential increase of tool wear. Under moderate milling conditions, the tool wear profile is similar to a cubic function. If the cutting conditions are aggressive, the slope of the steady state region increases, which means that the tool wear increase will be more like an exponential increase (Schmitz and Smith, 2008). This is reflected in the tool wear model developed using ALAMO. In the ALAMO model shown in Table 6, the log of tool wear is a function of the cutting time and the cube of the cutting time. Indeed, tool wear is an exponential function of the cutting time, evident from the tool wear vs Run# profile shown in Fig. 2. The exponent of the cube of the cutting time gives the model the nonlinear profile, that captures the slow increase in the tool wear followed by an exponential increase. The tool wear model was also a function of the product of spindle speed, indicator, and cutting time. This is interesting because spindle speed corresponds to the number of revolutions per minute by the tool, and the product of the spindle speed and cutting time gives the total number of revolutions performed by the tool for machining. This product is proportional to the cutting distance or the work done by the tool during machining. Therefore, it captures tool wear as a function of the cutting conditions and the time the tool has been machined. There are other terms in the ALAMO model, such as $\exp(s_i)$, v_s^2, and t_{cut}^2. ALAMO adds these terms to improve the fit of the tool wear model. ALAMO did not select the custom basis functions provided for model building because the width of cut, and depth of cut values in the data sets were constant. With a more diverse data, the ALAMO model would select the custom basis functions. However, this tool wear model is physics-informed and was able to represent how the tool wear progresses during machining, and building a model with more data would help build a model that considers the custom basis functions.

Table 4: ALAMO toolbox settings for model building.

Basis functions	Expressions: exp, log, ln , +, −, ×, \, constant Monomial power: 2, 3 Multi2Power: 1, 2, 3 Multi3Power: 1
Miscellaneous options	Fitness metric: BIC Screener: Lasso

The model obtained using GPTIPS is shown in the second row of Table 6. The tool wear model was a function of s_i, v_s^2, t_{cut}^3, $0.23 s_i t_{cut}$, and $t_{cut}^2 s_i$. As discussed for the tool wear model obtained using ALAMO, the tool wear was an exponential function of the cube of cutting time. This basis function captures the progression of tool wear during machining. The GPTIPS model discovered a relationship between tool wear and cutting time. GPTIPS selected the custom basis function of the product of three terms and also the function $0.23 s_i t_{cut}$. The tool wear model obtained by GPTIPS was an excellent fit of the experimental data, as shown in Fig. 2

Table 5: GPTIPS toolbox settings for model building.

Basis functions	Expressions: exp, ln, x^3, +, ×, \
Parameters	Population size: 100 Number of the generation: 100 Tournament size: 2 Max. tree depth : 5 Max. mutate depth : 2 Crossover (%): 84 (default) Mutation (%): 14 (default) Elitism (%): 15 (default) Direct reproduction (%): 2 (default)

Overall, the models obtained using ALAMO and GPTIPS provided a good estimate of the tool wear. Both models fit the data well, as shown in Fig. 2. The functions developed during model building are explainable by the physics of the process. This gives confidence in the structure of these models, and if more diverse data sets were available, ALAMO and GPTIPS would evolve better informed models that may make use of the custom basis functions inspired by the domain expert knowledge.

Table 6: Tool wear models obtained using ALAMO and GPTIPS for HAAS and DMG machines.

	Model	R^2
ALAMO	$\ln(W) = 3.77 t_{cut} - 1.93 \exp(s_i)$ $-0.32 v_s^2 - 5.32 t_{cut}^2 + 4.32 t_{cut}^3$ $+1.89 v_s s_i t_{cut}$	0.962
GPTIPS	$\ln(W) = -4.022 s_i + 0.68 v_s^2$ $+8.81 t_{cut}^3 - 3.55(0.23 s_i t_{cut})$ $-12.43(t_{cut}^2 s_i) - 2.09$	0.969

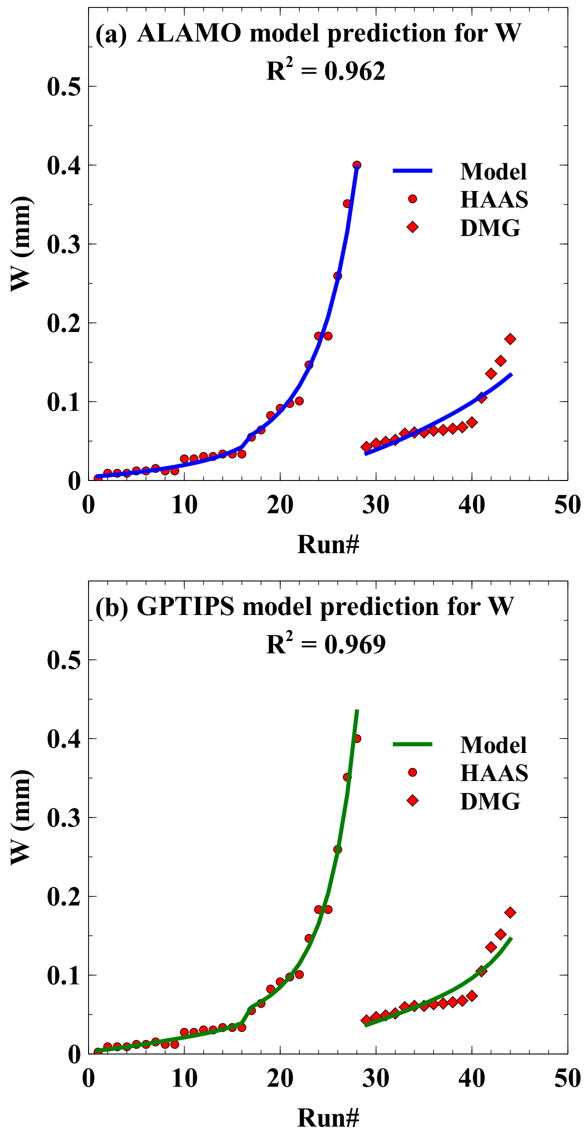

Figure 2: (a) Tool wear predictions of the surrogate model developed using ALAMO for tool wear data from experiments performed on HAAS and DMG machines, (b) tool wear predictions of the surrogate model developed using GPTIPS for tool wear data from experiments performed on HAAS and DMG machines.

Conclusion

Grey-box models are a better approach for building data-driven models that capture the physics of the system. These models are less complex in structure and are easy to build. This study discussed two approaches to building grey-box models using symbolic regression. ALAMO and GPTIPS, were used to build physics-informed surrogate models of tool wear for a manufacturing process. The key features of these models were that they were in an algebraic form of the input variables, which made the models interpretable. Both ALAMO and GPTIPS were shown successful in building surrogate models for complex physical phenomena for which the exact models do not exist. In future work, a generic and better surrogate models can be developed for tool wear using more diverse data.

Acknowledgement

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Advanced Manufacturing Office Award Number DE-EE0007613. We also gratefully acknowledge the Air Force Research Laboratory, Materials and Manufacturing Directorate (AFRL/RXMS) for support via Contract No. FA8650-20-C-5206

References

Awasthi, U. and Bollas, G. M. (2020). Sensor network design for smart manufacturing – Application on precision machining. *IFAC PapersOnLine*, 53(2):11440–11445.

Awasthi, U., Wang, Z., Mannan, N., Pattipati, K. R., and Bollas, G. M. (2022). Physics-based modeling and information-theoretic sensor and settings selection for tool wear detection in precision machining. *Journal of Manufacturing Processes*, 81(December 2021):127–140.

Binder, M., Klocke, F., and Doebbeler, B. (2017). An advanced numerical approach on tool wear simulation for tool and process design in metal cutting. *Simulation Modelling Practice and Theory*, 70:65–82.

Cozad, A., Sahinidis, N. V., and Miller, D. C. (2014). Learning surrogate models for simulation-based optimization. *AIChE Journal*, 60(6):2211–2227.

Cozad, A., Sahinidis, N. V., and Miller, D. C. (2015). A combined first-principles and data-driven approach to model building. *Computers and Chemical Engineering*, 73:116–127.

Koza, J. R. (1994). Genetic programming as a means for programming computers by natural selection. *Statistics and Computing*, 4(2):87–112.

Sansana, J., Joswiak, M. N., Castillo, I., Wang, Z., Rendall, R., Chiang, L. H., and Reis, M. S. (2021). Recent trends on hybrid modeling for industry 4.0. *Computers and Chemical Engineering*, 151:107365.

Schmidt, M. and Lipson, H. (2009). Distilling Natural Laws. *Science*, 324(April):81–85.

Schmitz, T. L. and Smith, K. S. (2008). *Machining Dynamics: Frequency Response to Improved Productivity*. Springer US.

Searson, D. P. (2015). *GPTIPS 2: An Open-Source Software Platform for Symbolic Data Mining*, pages 551–573. Springer International Publishing, Cham, Denmark.

Searson, D. P., Leahy, D. E., and Willis, M. J. (2010). GPTIPS: An Open Source Genetic Programming Toolbox For Multigene Symbolic Regression. In *Proceedings of the International Multi-Conference of Engineers and Computer Scientists*, Hong Kong.

Udrescu, S. M. and Tegmark, M. (2020). AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16).

Wilson, Z. and Sahinidis, N. (2017). ALAMO: Machine learning from data and first principles. *Process Development Symposium 2018: Applying New Technologies in Process Development*, pages 38–48.