Proceedings of the ASME 2015 International Manufacturing Science and Engineering Conference
MSEC2015
July 8-12, 2015, Charlotte, North Carolina, USA

MSEC2015-9354

A generalized data-driven energy prediction model with uncertainty for a milling machine tool using Gaussian Process

Jinkyoo Park and Kincho H. Law
Engineering Informatics Group
Stanford University
Stanford, CA, USA

Raunak Bhinge, Nishant Biswas, Amrita Srinivasan and David A. Dornfeld
Laboratory for Manufacturing and Sustainability
University of California
Berkeley, CA, USA

Moneer Helu and Sudarsan Rachuri
Systems Integration Division
National Institute of Standards and Technology
Gaithersburg, MD, USA

ABSTRACT
Using a machine learning approach, this study investigates the effects of machining parameters on the energy consumption of a milling machine tool, which would allow selection of optimal operational strategies to machine a part with minimum energy. Data-driven prediction models, built upon a nonlinear regression approach, can be used to gain an understanding of the effects of machining parameters on energy consumption. In this study, we use the Gaussian Process to construct the energy prediction model for a computer numerical control (CNC) milling machine tool. Energy prediction models for different machining operations are constructed based on collected data. With the collected data sets, optimum input features for model selection are identified. We demonstrate how the energy prediction models can be used to compare the energy consumption for the different operations and to estimate the total energy usage for machining a generic part. We also present an uncertainty analysis to develop confidence bounds for the prediction model and to provide insight into the vast parameter space and training required to improve the accuracy of the model. Generic parts are machined to test and validate the prediction model constructed using the Gaussian Process and we consistently achieve an accuracy of over 95% on the total predicted energy.

KEYWORDS
Energy prediction, Big data, Data-driven manufacturing, Gaussian Process

INTRODUCTION

The ability to accurately predict energy consumption during machining can provide several advantages. For example, energy prediction may be used to estimate energy costs and improve the efficiency of machine tools. It is also useful to predict energy consumption in response to new regulations and business drivers, such as Smart Grid and carbon cap-and-trade. Finally, energy prediction

can enable process monitoring since changes in power demand and energy consumption of machine tools can identify different events and states [1]. Despite their potential in managing and improving machine tool performance, data-driven energy prediction models remain scarce due to the difficulty in systematically measuring energy usage with all machining parameters and input features. However, recent technologies and standards have made it easier to efficiently monitor and manage the machining operation data needed to develop energy prediction models. One such example is MTConnect, which is a standard that has been developed to facilitate archiving, accessing, and retrieving operational data from manufacturing equipment [2]. MTConnect allows for the collection of raw sensor data and machine operational information, which provides a means to track variations in energy consumption by the different machining operations [1].

Machine tool monitoring for tool condition and energy consumption has been a subject of active research [3-6]. Advancements in machine automation and sensing technology have allowed measurement of the conditions and energy consumption of an operating machine. Teti *et al.* [7] gave an extensive survey of sensor technologies, signal processing, and decision-making methodologies on machine monitoring. Machine learning techniques, such as Artificial Neural Network [8-12], Support Vector Regression [13, 14], Hidden Markov Model [15, 16], and Conditional Random Field [17], have been proposed to study the relationships between machining operations and tool conditions. MTConnect has been employed to study the effects of different machining process parameters on the energy consumed by a machine tool to produce a part [18]; this work also constructed a statistical regression model for energy consumption. While prior studies have illustrated the potential of collecting operational and energy consumption data for further analysis, most research dealt primarily with a single operation (e.g., face milling). But, machining a part in practice involves multiple machining operations with different combinations of operational parameters. This study aims to develop energy prediction models for the various operations of a computer numerical control (CNC) milling machine using a machine learning approach. The energy prediction models are then aggregated and applied to estimate the total energy consumption for part machining.

EXPERIMENTAL DESIGN AND DATA PREPARATION

With rapid advancements in sensing and data management technologies, energy consumption data corresponding to each code block of an NC code and its corresponding machining parameters can be collected in real time and efficiently processed and retrieved remotely. This motivated us to design novel experiments to collect data that can realistically represent the manufacturing process by a machine tool. The experimental design, setup, and data processing techniques used for generating the training data for this study have been described previously by Bhinge *et al*. [19]. Here, we briefly present the basic setup and data processing steps used in our experiments. We then discuss the input features and data sets employed in constructing the prediction model.

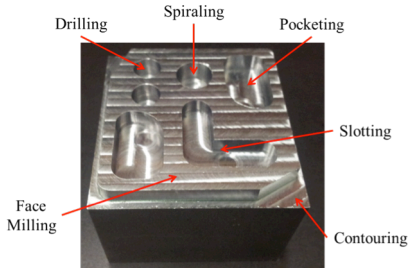


Figure 1. Test part for experimentation [19]

Table 1. Experimental setup [19]

Work piece Material	Cold Finish Mild Steel 1018
Work piece Dimensions	63.5mm x 63.5mm x 56mm
Machine Make	Mori Seiki NVD 1500
Machine Type	Micro NC Milling Machine
Tool Material	Solid Carbide
Tool Diameter	3/8" (9.525 mm)

Figure 1 shows a sample part designed to collect the training data for the data-driven energy prediction function [19]. Table 1 shows specific details of the work piece, the machine tool, and the cutting tool employed in the experimental study [19]. There are six basic cutting operations involved in machining the part: face milling, contouring, pocketing, slotting, spiraling and drilling. In addition, there are three non-cutting operations that are also included in the data sets collected from the experiments: air-cuts in $x-y$ and z directions and rapid tool motion. Because process parameters, such as feed rate, spindle speed, and depth of cut, can affect energy consumption, the test parts are produced with different combinations of machining parameters to investigate the relationship between the machining process parameters and the energy usage. A Taguchi technique [20] has been employed to design the experiments to ensure a fractional factorial combination for each of the process parameters in each operation. Table 2 shows the levels chosen for the depth of cut, chip load, and spindle speed used to machine the parts [19]. The feed rate f (mm/min) is obtained as the product of the spindle speed (RPM), the number of tool teeth, and the chip load (mm/tooth).

18 parts were machined for a total of 196 face milling experiments, 108 contouring experiments, 54 slotting and

Table 2. Levels for obtaining training data [19]

Level	Spindle Speed (RPM)	Chip Load (mm/tooth)	Depth of Cut (mm)
1	1500	0.0254	1
2	3000	0.0330	1.5
3	4500	0.0432	3
4	6000	0.0508	-

pocketing experiments, 16 spiraling experiments, and 32 drilling experiments. The face-milling operations on the first 9 parts were carried out in the y direction, while the remaining 9 parts were milled in the x direction. The separation of milling operations in the x- and y-directions was necessary to measure the energy consumption accurately for the target machine. The data sets collected during machining of the 18 parts were then used to construct the energy prediction model for each cutting or non-cutting operation.

Figure 2 shows the hardware used to collect the energy consumption data contextualized with machining parameters [21]. The machining process data, such as process parameters, NC blocks, and tool positions, were collected from the FANUC controller, and the power time series data was collected using a High Speed Power Meter from System Insights. The machining parameters and the power time series were synchronously combined through the MTConnect agent. The hardware platform and the data acquisition system used here were identical to the ones presented by Bhinge *et al*. [19] and Helu *et al*. [21].

The raw data collected from the MTConnect agent included the timestamp, power consumption, feed rate, spindle speed, and NC code block information. The collected data were then processed in two stages. The first stage calculated the derived data, such as the average feed rate, average spindle speed, and cumulative energy consumption for each NC code block. The second stage calculated the volume of material removed, depth of cut, and cutting strategy for each NC code block by simulating the cutting process while accounting for the actual dimensions of the work piece. The data obtained from the cutting simulation was denoted as simulated data. Figure 3 summarizes the data categorized as directly measured data, derived data, and simulated data, which were used to construct the energy prediction function for the machine tool [19].

From the NC code, the measured, derived, and simulated data were obtained as the input parameters for the learning algorithm. Specifically, the input feature vector $\boldsymbol{x} = \{x_1, \dots, x_5\}$ is defined as follows:

- $x_1 \in \mathbb{R}$ Feed rate: the velocity at which the tool is fed, which can be retrieved from the controller data
- $x_2 \in \mathbb{R}$ Spindle speed: rotational speed of the tool, which can be retrieved from the controller data
- $x_3 \in \mathbb{R}$ Depth of cut: the actual depth of material that the tool is removing, which can be obtained from the cutting simulation
- $x_4 \in \{1, 2, 3, 4\}$ Active tool axis ID (1 is for x-axis, 2 for y-axis 3 for z-axis and 4 for x-y axes): index for the active cutting direction, which can be determined by the lengths of cut in the x-, y- and z-directions
- $x_5 \in \{1, 2, 3\}$ Cutting strategy (1 is for conventional, 2 for climbing and 3 for both): the method for removing material, as obtained from the cutting simulation

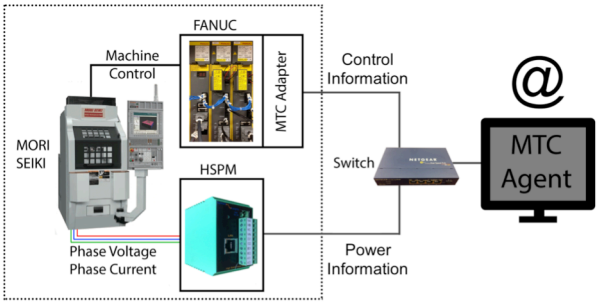


Figure 2. Data collection and processing hardware [21]

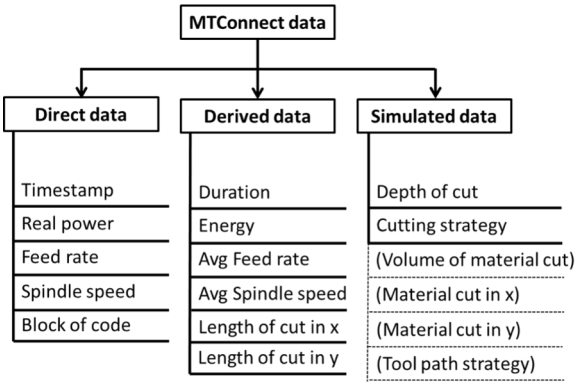
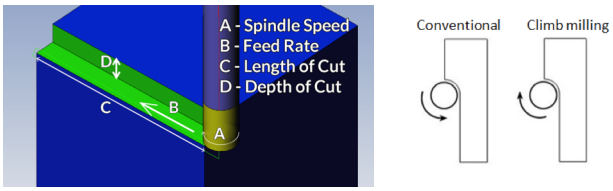


Figure 3. Classified data types from MTConnect [19]

The input process parameters used in this study are shown in Figure 4 [19]. Additionally, the total length of the tool path, $l \in \mathbb{R}$, in a single NC code block was computed using the lengths of cut in the x-, y- and z-directions. For the output response, the energy consumption of the milling machine measured by a high-speed power meter was used. Specifically, the average energy consumption, $E \in \mathbb{R}$, was obtained by numerically integrating the power time series measured during the machining operation in each NC code block. In this study, the output feature y was the energy density value defined as the energy consumption per unit length of tool path for each NC code block.



(a) Process parameters of a milling process (b) Cutting strategy

Figure 4. Machining process parameters [19]

In total, 12,299 pairs (NC code blocks) of input feature vectors $\boldsymbol{x}$ and output feature y were collected from the experiments. To ensure the training time series data were statistically of good quality, we selectively used the input features and the energy densities corresponding to NC code blocks whose duration were longer than 3 seconds (for rapid motion, time filtering was not applied since the operational durations were all under 3 seconds). The filtered data set $\boldsymbol{D} = \{(\boldsymbol{x}^i, y^i) | i = 1, \dots, m\}$, where $m = 3{,}092$, was further categorized into nine different data sets represented by $\{\boldsymbol{D}_1, \dots, \boldsymbol{D}_q, \dots, \boldsymbol{D}_9\}$ corresponding to the nine operations where each data set $\boldsymbol{D}_q = \{(\boldsymbol{x}^i, y^i) | i = 1, \dots, m_q\}$ contains m_q NC code blocks for the machining operation type q.

ENERGY PREDICTION MODEL

This section discusses how the energy prediction models for the machining operations are constructed using a Gaussian Process (GP). First, the basic procedure for constructing a GP regression model for a machining operation is introduced. We then discuss the selection of optimum input features that provide the best prediction accuracy for a machining operation. Lastly, we describe how the prediction models for each machining operation are aggregated and applied to estimate the total energy consumption for machining a part.

Gaussian Process Regression

In the GP, we assume that the output $y = f_q(\boldsymbol{x}) + \epsilon$ is measured with noise $\epsilon \sim N(0, \sigma_\epsilon^2)$, which is also Gaussian distributed with zero mean and variance σ_ϵ^2. In a GP, the values for the unknown function $f_q(\boldsymbol{x})$ are treated as random variables and modeled by a Gaussian distribution for incorporating prior knowledge captured in the historical data. Suppose the current data set is denoted by $\boldsymbol{D}_q = \{(\boldsymbol{x}^i, y^i) | i = 1, \dots, m_q\}$ for the machining operation type q. The measured output $y^{new} = f_q(\boldsymbol{x}^{new}) + \epsilon^{new}$ corresponding to the new input feature $\boldsymbol{x}^{new}$ and the historical outputs $\boldsymbol{y}^{1:m_q} = \{y^1, \dots, y^{m_q}\}^T$ in the training data set $\boldsymbol{D}_q$ follow a multivariate Gaussian distribution [22]:

$$\begin{bmatrix} \boldsymbol{y}^{1:m_q} \\ y^{new} \end{bmatrix} \sim N\left(\boldsymbol{0}, \begin{bmatrix} \mathbf{K} & \boldsymbol{k} \\ \boldsymbol{k}^T & k(\boldsymbol{x}^{new}, \boldsymbol{x}^{new}) \end{bmatrix}\right), \quad (1)$$

where $\boldsymbol{k}^T = \{k(\boldsymbol{x}^1, \boldsymbol{x}^{new}), \dots, k(\boldsymbol{x}^{m_q}, \boldsymbol{x}^{new})\}$ and $\mathbf{K}$ is the covariance matrix (kernel matrix) whose (i, j)th entry is $\mathbf{K}_{ij} = k(\boldsymbol{x}^i, \boldsymbol{x}^j)$. The value of the covariance function $k(\boldsymbol{x}^i, \boldsymbol{x}^j)$ quantifies the amount the two input feature vectors $\boldsymbol{x}^i$ and $\boldsymbol{x}^j$ change together. Note that the more the two vectors $\boldsymbol{x}^i$ and $\boldsymbol{x}^j$ disagree, the closer the value of the covariance approaches zero, implying that the two input vectors are not correlated in terms of their function values. We use a squared exponential function to evaluate the covariance between the two input feature vectors $\boldsymbol{x}^i$ and $\boldsymbol{x}^j$ as [23]:

$$k(\boldsymbol{x}^i, \boldsymbol{x}^j) = \tau^2 \exp\left(-\frac{1}{2}(\boldsymbol{x}^i - \boldsymbol{x}^j)^T \operatorname{diag}(\boldsymbol{\lambda})^{-2}(\boldsymbol{x}^i - \boldsymbol{x}^j)\right) + \sigma_\epsilon^2 \delta_{ij}. \quad (2)$$

In Eq. (2), the Kronecker delta function δ_{ij} serves to selectively specify the noise variance σ_ϵ^2 to the covariance value; that is, the noise signals adding to different measurements are assumed to be independent and the noise correlation is non-zero only when $i = j$. The covariance function is described by the hyper parameters τ, which denotes the amplitude of the function, and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_i, \dots \lambda_5)$, where the length scale λ_i quantifies the relevancy of the i^{th} input feature x_i (i =1, …,5) in $\boldsymbol{x}$ for predicting the energy consumption. A large length scale indicates weak relevance, while a small length scale implies strong relevance of the corresponding machining parameter in predicting the energy consumption. The optimum hyper parameters τ and $\boldsymbol{\lambda}$ can be obtained by maximizing the log-likelihood of data [22]. In this study, we use a machine-learning module, scikit-learn [24], to construct the GP regression models. Once the length scales $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_i, \dots \lambda_5)$ are determined, the importance of each input feature x_i in predicting output y for the specific machining operation can also be studied.

Since the distribution conditional on any subset of the data assumed to be Gaussian distributed is also Gaussian, the posterior distribution $p(y^{new} | \boldsymbol{D}_q, \boldsymbol{x}^{new})$ on y^{new} given the historical data set $\boldsymbol{D}_q = \{(\boldsymbol{x}^i, y^i) | i = 1, \dots, m_q\}$ and the new input feature vector $\boldsymbol{x}^{new}$ can be expressed as a 1-D Gaussian distribution [22]:

$$p(y^{new} | \boldsymbol{D}_q, \boldsymbol{x}^{new}) = N\left(y^{new}; \mu(\boldsymbol{x}^{new} | \boldsymbol{D}_q), \sigma^2(\boldsymbol{x}^{new} | \boldsymbol{D}_q)\right). \quad (3)$$

The posterior distribution $p(y^{new} | \boldsymbol{D}_q, \boldsymbol{x}^{new})$ can be described by its mean μ and the variance σ^2, which can be expressed, respectively, as [22]:

$$\mu(\boldsymbol{x}^{new} | \boldsymbol{D}_q) = \boldsymbol{k}^T \mathbf{K}^{-1} \boldsymbol{y}^{1:m_q}, \quad (4)$$

$$\sigma(\boldsymbol{x}^{new}|\boldsymbol{D}_q) = \sqrt{k(\boldsymbol{x}^{new}, \boldsymbol{x}^{new}) - \boldsymbol{k}^T\mathbf{K}^{-1}\boldsymbol{k}}\,. \quad (5)$$

That is, we can obtain the mean function $\mu(\boldsymbol{x}|\boldsymbol{D}_q)$ from the Gaussian Process to predict the most probable energy density $\hat{y} = f_q(\boldsymbol{x})$ for a given input feature vector $\boldsymbol{x}$ and the standard deviation function $\sigma(\boldsymbol{x}|\boldsymbol{D}_q)$ to quantify the uncertainty in the predicted value of $\hat{y}$ at $\boldsymbol{x}$. As will be discussed, the energy consumption per each machining operation is aggregated to predict the total energy consumption (with some estimated uncertainty bound) for machining a part.

Optimum Input Feature Selection

Depending on the machining operation, the parameters in the input feature vector $\boldsymbol{x}$ affect the energy density value y differently. For each machining operation type q with the data set $\boldsymbol{D}_q = \{(\boldsymbol{x}^i, y^i)|i = 1, \dots, m_q\}$, we select the optimum combination of the input features from $\boldsymbol{x} = \{x_1, x_2, x_3, x_4, x_5\}$ that best predicts the energy density. Only the selected input features will then be used to model the energy density prediction function. The optimum input feature sets are selected using the holdout cross validation technique [25].

First, we select one possible combination of input features from a set of candidate input features among all possible (factorial) combinations of the candidate input features. For each combination of input features, we compute the error rates from the GP models as follows:

(1) Randomly divide the data set $\boldsymbol{D}_q$ into the training data set $\boldsymbol{D}_q^{tr}$ with m_q^{tr} training data points and the validating data set $\boldsymbol{D}_q^{va}$ with m_q^{va} validation data points. In this study, we set the ratio $m_q^{tr}: m_q^{va} = 7:3$.

(2) Construct the energy density prediction function $f_q(\boldsymbol{x})$ by computing $\mu(\boldsymbol{x}|\boldsymbol{D}_q)$ and $\sigma(\boldsymbol{x}|\boldsymbol{D}_q)$ using the training data set $\boldsymbol{D}_q^{tr}$.

(3) Predict the energy densities corresponding to the input features in the validation data set $\boldsymbol{D}_q^{va}$ and compute the error rates by comparing them to the true energy densities in the validation data set $\boldsymbol{D}_q^{va}$. The error is measured in terms of the mean absolute error (MAE), which is more insensitive to outliers than the root mean square error (RMSE) [26]:

$$\mathrm{MAE}_q = \frac{1}{m_q^{va}} \sum\nolimits_{\{(\boldsymbol{x}^i, y^i) \in \boldsymbol{D}_q^{va}\}} \left|\mu(\boldsymbol{x}^i|\boldsymbol{D}_q) - y^i\right|. \quad (6)$$

MAE_q quantifies how close the predictions are compared to the measured values. Specifically, we use the normalized mean absolute error (NMAE) [27]:

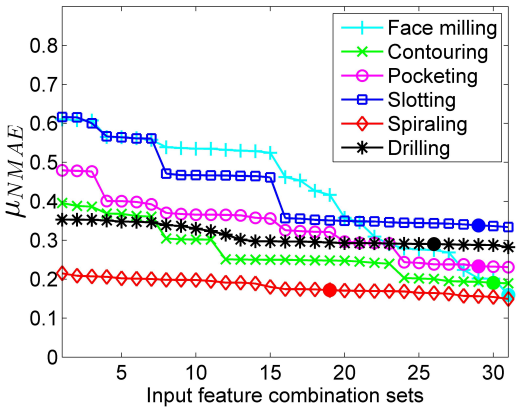
$$\mathrm{NMAE}_q = \frac{\sum_{\{(\boldsymbol{x}^i, y^i) \in \boldsymbol{D}_q^{va}\}} \left|\mu(\boldsymbol{x}^i|\boldsymbol{D}_q) - y^i\right|}{\sum_{\{(\boldsymbol{x}^i, y^i) \in \boldsymbol{D}_q^{va}\}} y^i} = \frac{\mathrm{MAE}_q}{\bar{y}_q}. \quad (7)$$

Note that we can compute the average deviation between the predicted and measured densities, *i.e.*, MAE_q, by simply multiplying NMAE_q with the measured mean density $\bar{y}_q$ (for the machining operation type q).

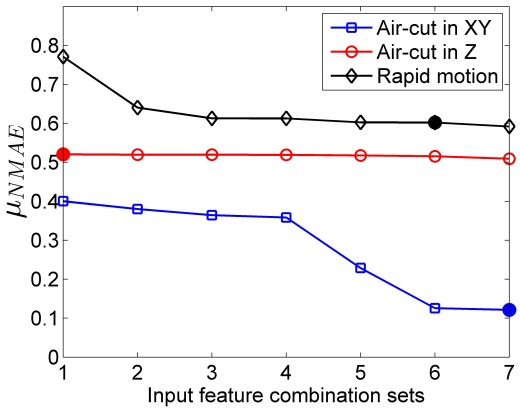
The above procedure is repeated 100 times in this study, and the averaged error rate μ_{NMAE} is computed to minimize the dependency of the error rate on specific training and validation sets. Finally, among all possible combinations of input features, the optimum input feature set that gives the lowest averaged error rate is selected. The selected optimum input feature set will then be used to construct the energy density prediction functions.

Figures 5(a) and 5(b) show the average NMAE, μ_{NMAE}, for the cutting and non-cutting operations. For the cutting operations, all input features, x_1 (feed rate), x_2 (spindle speed), x_3 (depth of cut), x_4 (active tool axis), and x_5 (cutting method), are included as candidates for selection with a total of 31 possible combinations of input feature sets. For the non-cutting operations, only the input features, x_1, x_2, and x_4 are considered because the input features x_3 (depth of cut) and x_5 (cutting method) are irrelevant for the operations. There are altogether 7 different combinations of input feature sets for the non-cutting operations. For each machining operation type, the μ_{NMAE} values for all the different combinations of input features are sorted and plotted in Figures 5(a) and 5(b). As marked in the figures, the solid circles indicate the average error rates for each machining operation when all possible input features are included. It can be seen that including all the available input features may result in over fitting and does not necessarily achieve the lowest error ratio.

Table 3 shows the optimum input features $\boldsymbol{x}^*$ for the machining operations and the corresponding average error rates μ_{NMAE}. From the optimum input features $\boldsymbol{x}^*$, we can also rank the relative importance of individual input feature x_i based on the residual accuracy, defined as $\mu_{\mathrm{NMAE}}(\boldsymbol{x}^*) - \mu_{\mathrm{NMAE}}(\boldsymbol{x}^* \setminus \{x_i\})$, where $\mu_{\mathrm{NMAE}}(\boldsymbol{x}^* \setminus \{x_i\})$ represents the average error rate with the input feature x_i removed from the optimum input feature vector $\boldsymbol{x}^*$. The larger the residual accuracy when the input feature x_i



(a) Cutting processes



(b) Non-cutting processes

Figure 5. Change in the NMAE of the GP for different sets of input feature combinations

Table 3: Optimum input feature sets for different processes
(x_1 : feed rate, x_2 : spindle speed, x_3 : depth cut, x_4 : cutting direction, x_5 : cutting method.)

	Operation type	Number of NC blocks	Average duration (sec)	Optimum input features	$\mu_{\bar{y}_q}$ (J/mm)	μ_{MAE} (J/mm)	μ_{NMAE} (%)
Feed with cut	Face milling	1,466	20.8	$\{x_4, x_1\, x_5, x_3, x_2\}$	3.072	0.502	16.35
	Contouring	425	9.2	$\{x_4, x_1, x_3, x_2\}$	4.612	0.833	18.08
	Slotting	134	5.6	$\{x_4, x_1, x_2, x_5\}$	4.372	1.028	23.60
	Pocketing	168	5.4	$\{x_4, x_1, x_2\}$	3.873	1.306	33.74
	Spiraling	16	5.4	$\{x_1, x_4, x_3\}$	6.420	1.009	15.94
	Drilling	140	22.2	$\{x_1, x_2, x_4, x_3\}$	10.845	2.722	25.85
Non-cutting	Air cut in x-y	140	12.3	$\{x_1, x_2, x_4\}$	4.037	0.5314	13.27
	Air cut in z	281	6.5	$\{x_4, x_1, x_2\}$	10.336	4.796	46.27
	Rapid motion	322	0.4	$\{x_2, x_4\}$	0.558	0.323	58.04

is excluded implies that the input feature x_i contributes to better prediction accuracy. In Table 3, the input features for each machining operation are listed in the order of their relative importance. For example, in the case of face milling, the cutting direction x_4 is ranked as the most important machining parameter among the five input features. This result indicates that the machine consumes different levels of energy with the tool moving in different directions.

Uncertainty Estimation for Energy Prediction

With the energy density prediction model for each machining operation type q represented by the mean energy density function $\mu_q(\boldsymbol{x}|\boldsymbol{D}_q)$ and the associated standard deviation function $\sigma_q(\boldsymbol{x}|\boldsymbol{D}_q)$, the total energy consumption for machining a part can be estimated from the NC codes. First, from the input feature $\boldsymbol{x}^i$ of the NC code block i performing the machining operation type q, we can estimate the energy consumption $\hat{E}^i$ and the standard deviation S^i, respectively, as:

$$\hat{E}^i = \mu_q(\boldsymbol{x}^i|\boldsymbol{D}_q)\times l^i, \tag{8}$$

$$S^i = \sigma_q(\boldsymbol{x}^i|\boldsymbol{D}_q)\times l^i, \tag{9}$$

where l^i is the length of tool path specified for the operation by the NC code. Aggregating all the NC blocks for the machining operation type q, the predicted total energy consumption $\hat{E}_q$ and the associated standard deviation S_q can be computed for that operation type:

$$\hat{E}_q = \sum_{\{(\boldsymbol{x}^i, y^i)\in \boldsymbol{D}_q\}} \mu_q(\boldsymbol{x}^i|\boldsymbol{D}_q)\times l^i, \tag{10}$$

$$S_q = \sqrt{\sum_{\{(\boldsymbol{x}^i, y^i)\in \boldsymbol{D}_q\}} \left(\sigma_q(\boldsymbol{x}^i|\boldsymbol{D}_q)\times l^i\right)^2}. \tag{11}$$

Finally, the estimated total energy consumption $\hat{E}$ for machining a whole part and the standard deviation S associated with the estimation can be computed, respectively, by summing the predicted energy $\hat{E}_q$ and accumulating the standard deviation S_q for all machining

operation types, $q = 1, \ldots, Q$, where Q = 9 (including all cutting and non-cutting operations), as

$$\hat{E} = \sum_{q=1}^{Q} \hat{E}_q, \tag{12}$$

$$S = \sqrt{\sum_{q=1}^{Q} (S_q)^2}. \tag{13}$$

Note that the energy density function $\hat{y}$ for an operation type q, represented by its mean function $\mu_q(\boldsymbol{x}|\boldsymbol{D}_q)$ and standard deviation function $\sigma_q(\boldsymbol{x}|\boldsymbol{D}_q)$, is assumed to be Gaussian, the predicted total energy $\hat{E}$ computed as a linear combination of the energy densities, as shown in Eqs. (10) and (12), is also Gaussian, according to the Gaussian Process. Given the predicted total energy $\hat{E}$ and the predicted total standard deviation S, the probability density function for the true total energy consumption E can be represented as $E \sim N(\hat{E}, S)$.

ENERGY PREDICTION OF GENERIC PARTS

We apply the energy prediction and uncertainty estimation functions constructed from the training data sets to predict the energy consumption for machining a generic part to assess the accuracy of the GP. Figure 6 shows the generic part, the geometry of which is quite different from the part used in the training process. The machining operations used to produce the generic test part involve the cutting operations face milling, pocketing, and drilling and the non-cutting operations air-cut in x-y direction, air cut in z direction, and auxiliary operations (such as rapid tool motion).

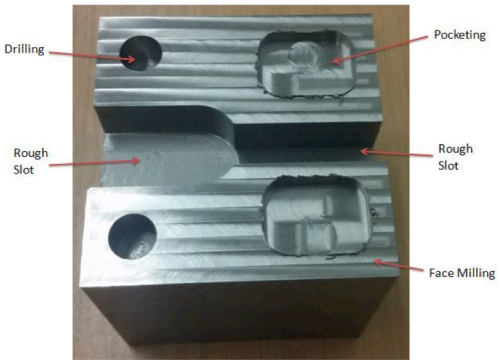


Figure 6. Generic test part

Table 4: Spindle speeds chosen for the blind tests

	Used spindle speeds (in RPM)
Training parts 1~18	{1,500; 3,000; 4,500}
Test part 1	{1,500; 3,000; 4,500}
Test part 2	{1,700; 2,800; 4,300}
Test part 3	{2,125; 2,400; 3,750}

Three test parts with identical geometry are machined but using different sets of spindle speeds, as shown in Table 4. Comparing the spindle speeds used in machining the 18 training parts, the first test part uses the same spindle speed while the second and third use different spindle speeds from those of the training sets. For all test parts, the depth of cut is set to 1 mm.

Figure 7 shows the measured energy density values y and the prediction energy density function $\hat{y}$ for the face milling operations with different spindle speeds and different feed rates. For the prediction energy density function, the plots show a one-standard deviation bound, i.e., $\mu_1(\boldsymbol{x}^i|\boldsymbol{D}_1) \pm \sigma_1(\boldsymbol{x}^i|\boldsymbol{D}_1)$. As shown in Figure 7(a), the measurements from test part 1 are well captured by the prediction functions. For test parts 2 and 3, the measurements are also well captured by the prediction function shown in Figures 7(b) and 7(c). However, underestimations by the prediction models on the energy density are observed when the spindle speeds are high.

Figure 8 compares the mean predicted and measured energy consumption for each individual NC code block. In general, the mean predicted energy consumption values match well with the measurements. However, differences can be observed in test parts 2 and 3 in the NC block sequence between 95 and 140 when the spindle speeds are high. Finally, Table 5 compares the total mean predicted and measured energy consumption using NMAE (normalized mean absolute error) and RTE (relative total error) defined as:

$$\text{NMAE} = \frac{\sum_{\{i \in \text{ NC blocks}\}} \left|\hat{E}^i - E^i\right|}{\sum_{\{i \in \text{ NC blocks }\}} E^i}, \tag{14}$$

$$\text{RTE} = \frac{\left|\hat{E} - E\right|}{E}. \tag{15}$$

Note that the NMAE in Eq. (14) is defined using the predicted energy $\hat{E}^i$ and the measured energy E^i, whereas NMAE in Eq. (7) is defined using the predicted density $\hat{y}^i$ and the measured energy density y^i. Thus, the energy prediction with the longer length of tool path l^i will contribute more to the value of NMAE in Eq. (14). In spite of this dependence on the geometry, the measure can still quantify the mean absolute errors of the three test cases (parts) in a relative manner.

It can be seen that for all three test cases, the relative total errors for predicting the total energy consumption for machining the three parts fall within 4%. The RTEs for the energy prediction are small for all three test parts because the distribution of errors $\hat{E}^i - E^i$ are centered at the zero-mean with almost equal chance to over- or under-estimate the energy as shown in Figure 9.

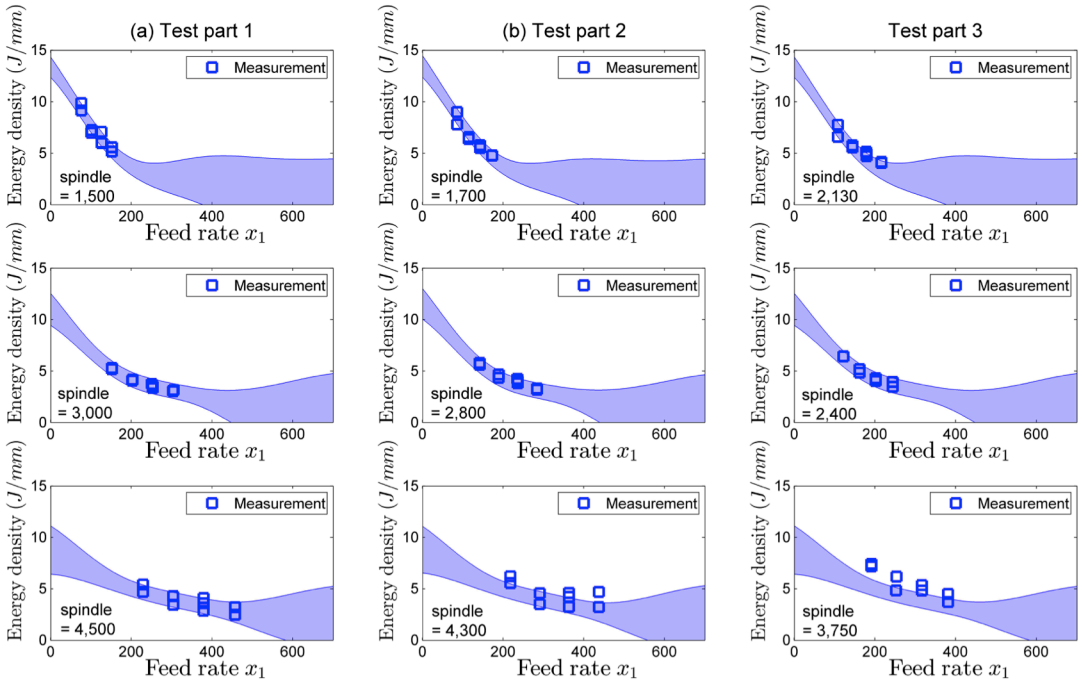
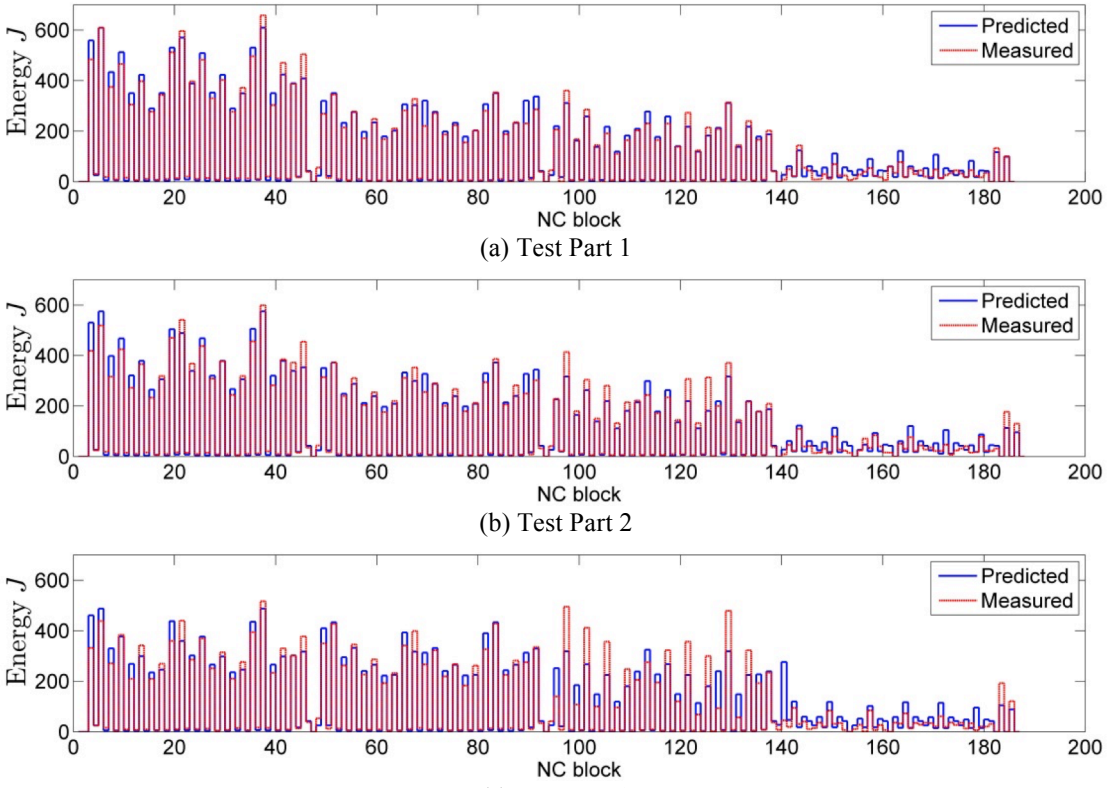


Figure 7. Predicted energy density values for generic test parts (machined using face-milling, y-direction cut, conventional cutting strategy and depth of cut = 1mm); the highlighted area shows the one standard deviation bounds for the prediction



(a) Test Part 1

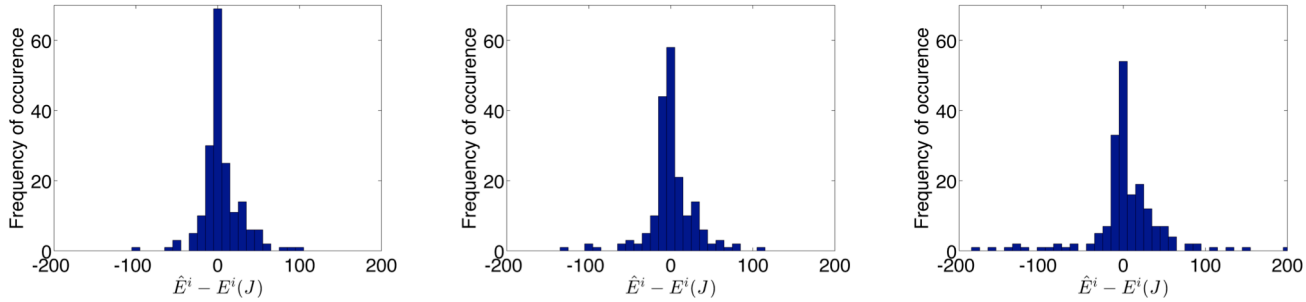
(b) Test Part 2

(c) Test Part 3

Figure 8. Total mean predicted energy consumption including all operations

Table 5: Summary of prediction results on the generic test parts

	No. of data	Averaged block duration (sec)	NMAE (%)	Measured Total energy (KJ)	Prediction total energy (KJ)	Standard deviation (KJ)	RTE (%)
Test 1	188	10.27	13.004	22.492	21.689	0.434	3.702
Test 2	188	9.82	15.210	21.928	21.864	0.441	0.290
Test 3	188	9.70	23.143	21.746	21.074	0.477	3.192



(a) Test part 1 (b) Test part 2 (c) Test part 3

Figure 9. Distributions of errors

DISCUSSION

This study demonstrates the use of a non-parametric regression model, namely the Gaussian Process (GP), to predict the energy consumption of a machine tool. The GP models the complex relationship between the input machining parameters and output energy consumption and constructs a prediction function for the energy consumption with confidence bounds. Using the GP, this study also examines the input features used in each prediction function and the relative importance of the input features for the machining operations. The probabilistic model also provides confidence bounds for the predictive estimations. Even though the training data sets in this study include only 18 experimental parts, the models constructed using the machine learning approach are able to accurately predict the energy consumption for machining a generic test part with the milling machine tool. The prediction models can be updated and, possibly, improved as further experimental data sets are collected.

The GP can potentially be extended as a real-time adaptive learning algorithm to reduce the burden of extensive training data collection. The technique can be used to develop the prediction function as and when new parameter spaces are explored. This, in turn, can significantly reduce data processing time, data storage capacity, and the dependence on using large training data sets. We are currently investigating a real-time adaptive system for energy and machine tool monitoring.

In conclusion, this study shows that with advanced data collection and processing techniques, prediction models can be constructed to predict energy consumption of a machine tool with multiple operations and multiple process parameters. The prediction model can be generalized to study energy consumption of a generic part with a degree of uncertainty based on the parameters employed to construct the model.

ACKNOWLEGEMENT AND DISCLAIMER

The authors acknowledge the support of the Smart Manufacturing Systems Design and Analysis Program at the National Institute of Standards and Technology (NIST), Grant Numbers 70NANB12H225 and 70NANB12H273 awarded to University of California, Berkeley and Stanford University, respectively. In addition, the authors appreciate the support of the Machine Tool Technologies Research Foundation (MTTRF) and System Insights for the equipment used in this research. Certain commercial systems are identified in this paper. Such identification does not imply recommendation or endorsement by NIST; nor does it imply that the products identified are necessarily the best available for the purpose.

REFERENCES

[1] Vijayaraghavan, A., and Dornfeld, D. (2010). "Automated energy monitoring of machine tools," *CIRP Annals – Manufacturing Technology*, 59, pp. 21-24.

[2] MTConnect Institute (2014). MTConnect v. 1.3.0, http://www.mtconnect.org/downloads/standard.aspx.

[3] Neugebauer, R., Denkena, B., and Wegener, K. (2007). "Mechatronics Systems for Machine Tools," *Annals of the CIRP,* 56, pp. 657-686.

[4] Dietmair, A., Verl, A., and Eberspaecher, P. (2009). "Predictive Simulation for Model Based Energy

Consumption Optimization in Manufacturing Systems Machine Control," *19th International Conference on Flexible Automation and Intelligent Manufacturing,* Middleborough, UK.

[5] Draganescu, F., Gheorghe, M., and Doicin, C.V. (2003). "Models of Machine Tool Efficiency and Specificific Consumed Energy," *Journal of Material Processing Technology,* 141, pp. 9-15.

[6] Diaz, N., Helu, M., Jarvis, A., Tonissen, S., Dornfeld, D., and Schlosser, R (2009). "Strategies for minimum energy operation for precision machining," *proceeding of MTTRF 2009 Annual Meeting*, Shanghai, China.

[7] Teti, R., Jemielniak, K., O'Donnell, G., and Dornfeld, D. (2010). "Advanced monitoring of machine operations," *CIRP Annals-Manufacturing Technology*, 50, pp. 717-739.

[8] Asiltürk, I., and Cunkas, M. (2011) "Modeling and Prediction of Surface Roughness in Turning Operations Using Artificial Neural Network and Multiple Regression Method," *Expert Systems with Application*, 38, pp. 5826-5832.

[9] Özel, T., and Karpat, Y. (2005). "Predictive Modeling of Surface Roughness and Tool Wear in Hard Turning Using Regression and Neural Networks," *International Journal of Machine Tools and Manufacture*, 45, pp. 467-479

[10] Oktem, H., Erzurumlu, T., and Erzincanli, F. (2006). "Prediction of Minimum Surface Roughness in End Milling Mold Parts Using Neural Network and Genetic Algorithm," *Materials and Design*, 27, pp. 735-744.

[11] Benardos, P.G., and Vosniakos, G.C. (2002). "Prediction of Surface Roughness in CNC Face Milling Using Neural Networks and Taguchi's Design of Experiments," *Robotics and Computer Integrated Manufacturing*, 18, pp. 343-354.

[12] Zain, A., Haron, H, and Sharif, S. (2010). "Prediction of Surface Roughness in the End Milling Machining Using Artificial Neural Network," *Expert Systems with Applications*, 37, pp. 1755-1768.

[13] Bhattacharyya, P., and Sanadhya, S.K. (2006). "Support Vector Regression Based Tool Wear Assessment in Face Milling," *Proceeding of IEEE International Conference on Industrial Technology, Mumbai*, India.

[14] Shi, D., and Gindy, N. (2007). "Tool Wear Predictive Model Based on Least Squares Support Vector Machines," *Mechanical System and Signal Processing*, 21, 1799-1814.

[15] Wang, L., Mehrabi, M., and Kannatey-Asibu, E. (2002). "Hidden Markov Model-based Tool Wear Monitoring in Turning," *Journal of Manufacturing Science and Engineering*, 124, pp. 651-658.

[16] Zhu, K., Wong, Y., and Hong, J. (2009). "Multi-category Micro-milling Tool Wear Monitoring with Continuous Hidden Markov Models," *Mechanical Systems and Signal Processing*, 23, 547-560.

[17] Wang, G., and Feng, X. (2013). "Tool Wear State Recognition Based on Linear Chain Conditional Random Field Model," *Engineering Applications of Artificial Intelligence*, 26, pp. 1421-1427.

[18] Diaz, N., Redelsheimer, E., and Dornfeld, D. (2011). "Energy Consumption Characterization and Reduction Strategies for Milling Machine Too Use," *18th CIRP International Conference on Life Cycle Engineering,* Braunschweig, Germany.

[19] Bhinge, R., Park, J., Biswas, N., Helu, M., and Dornfeld, D., Law, K., and Rachuri, S. (2014). "An Intelligent Machine Monitoring System Using Gaussian Process Regression for Energy Prediction," *IEEE International Conference on Big Data* (IEEE BigData 2014), Washington, DC.

[20] Box, G., Hunter, J.S., and Hunter, W.G. (1979). "Statistics for Experimenters: Design, Inovation, and Discovery," Wiley.

[21] Helu, M., Robinson, S., Bhinge, R., Bänziger, T., and Dornfeld, D. (2014). "Development of a Machine Tool Platform to Support Data Mining and Statistical Modeling of Machining Processes." *Proc MTTRF 2014 Annual Meeting*, San Francisco, CA, 2014.

[22] Rasmussen, C., and Williams, C. (2006). "Gaussian Process for machine learning," MIT Press.

[23] Neal, R.M. (1996). "Bayesian learning for neural networks," New. York: Springer-Verlag.

[24] Pedregosa *et al*., "Scikit-learn: Machine Learning in Python," JMLR, 12, 2011, pp. 2825-2830.

[25] Hastie, T., Tibshirani, R., and Friedman, J. (2009) "The Elements of Statistical Learning," Springer.

[26] Willmott, C., and Matsuura, K. (2005). "Advantage of the Mean Absoute Error over the Root Mean Square Error (RMSE) in Assessing Average Model Performane," *Climate Research*, 30, pp. 79-82.

[27] Gustafson, W.I., and Shaocai, Y. (2012). "Generalized Approach for Using Unbiased Symmetric Metrics with Negative Values: Normalized Mean Bias Factor and Normalized Mean Absolute Error Factor," *Atmospheric Science Letter*, 13(4), pp. 262-267.